

Computer Science & Information Technology

173

Computer Science and Information Technology

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 10th International Conference on Signal, Image Processing and Pattern Recognition (SIPP 2022)
- 3rd International Conference on Natural Language Processing & Computational Linguistics (NLPCL 2022)
- 3rd International conference on Big Data, Machine learning and Applications (BIGML 2022)
- 7th International Conference on Software Engineering (SOEN 2022)
- 10th International Conference on Artificial Intelligence, Soft Computing (AISC 2022)
- 7th International Conference on Networks, Communications, Wireless and Mobile Computing (NCWMC 2022)
- 12th International Conference on Computer Science and Information Technology (CCSIT 2022)

Published By



AIRCC Publishing Corporation

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Dhinaharan Nagamalai (Eds),
Wireilla Net Solutions, Australia
E-mail: dhinthia@yahoo.com

ISSN: 2231 - 5403
ISBN: 978-1-925953-72-5
DOI: 10.5121/csit.2022.121301 - 10.5121/csit.2022.121320

This work is subject to copyright. All rights are reserved, whether whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the International Copyright Law and permission for use must always be obtained from Academy & Industry Research Collaboration Center. Violations are liable to prosecution under the International Copyright Law.

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

10th International Conference on Signal, Image Processing and Pattern Recognition (SIPP 2022), July 30~31, 2022, London, United Kingdom, 3rd International Conference on Natural Language Processing & Computational Linguistics (NLPCL 2022), 3rd International conference on Big Data, Machine learning and Applications (BIGML 2022), 7th International Conference on Software Engineering (SOEN 2022), 10th International Conference on Artificial Intelligence, Soft Computing (AISC 2022), 7th International Conference on Networks, Communications, Wireless and Mobile Computing (NCWMC 2022), 12th International Conference on Computer Science and Information Technology (CCSIT 2022) was collocated with 12th International Conference on Computer Science and Information Technology (CCSIT 2022). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The SIPP 2022, NLPCL 2022, BIGML 2022, SOEN 2022, AISC 2022, NCWMC 2022 and CCSIT 2022. Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically.

In closing, SIPP 2022, NLPCL 2022, BIGML 2022, SOEN 2022, AISC 2022, NCWMC 2022 and CCSIT 2022 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the SIPP 2022, NLPCL 2022, BIGML 2022, SOEN 2022, AISC 2022, NCWMC 2022 and CCSIT 2022

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld,
Dhinaharan Nagamalai (Eds)

General Chair

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Organization

Southeastern Louisiana University, USA
Wireilla Net Solutions, Australia

Program Committee Members

Aashish A.Bardekar,
Abdalhossein Rezai,
Abdel-Badeeh M. Salem,
Abdelhadi Assir,
Abdelhalim Kessal,
Abdellah Yousfi,
Abdellatif I. Moustafa,
Abderrahim Siam,
Abderrahmane EZ-Zahout,
Abdessamad Belangour,
Abdulhamit Subasi,
Abhishek Shukla,
Addisson Salazar,
AdrianOlaru,
Ahmad A. Saifan,
Akhil Gupta,
Alejandro Regalado Mendez,
Alessio Ishizaka,
Ali Abdrhman Mohammed Ukasha,
Alireza Valipour Baboli,
Amal Azeroual,
Amari Houda,
Amizah Malip,
Ana Luísa Varani Leal,
Anastasios Doulamis,
Andy Rachman,
Anirban Banik,
Anita Dixit,
Anita Yadav,
Anouar Abtoy,
Archit Yajnik,
Aridj Mohamed,
Ashwath Rao B,
Assem Abdel Hamied Moussa,
Atanu Nag,
Ayman Kassem,
Azah Kamilah Muda,
Azian Azamimi Abdullah,
Azka Kishwar,
B Nandini,
B.K. Tripathy,
Badir Hassan,
Bahaa K. Saleh,
Baihua Li,

Sipna College of Engineering & Technology, India
University of Science and Culture, Iran
Ain Shams University, Egypt
Hassan 1st University, Morocco
University of Bordj Bou Arreridj, Algeria
University Mohammed V, Morocco
Umm AL-Qura University, Saudi Arabia
University of Khenchela, Algeria
Mohammed V University, Morocco
University Hassan II Casablanca, Morocco
Effat University, Saudi Arabia
R D Engineering College, India
Universitat Politècnica de València, Spain
University Politehnica of Bucharest, Romania
Yarmouk university, Jordan
Lovely Professional University, India
Universidad del Mar, Mexico
NEOMA Business School, France
Sebha University, Libya
University Technical and Vocational, Iran
Mohammed V University, Morocco
Networking & Telecom Engineering, Tunisia
University of Malaya, Malaysia
University of Macau, China
National technical University of Athens, Greece
Institut Teknologi Adhi Tama Surabaya, Indonesia
National Institute of Technology Agartala, India
SDM College of Engineering Dharwad, India
Harcourt Butler Technical University, India
Abdelmalek Essaadi University, Morocco
Sikkim Manipal Institute of technology, India
Hassiba Benbouali University Chlef, Algeria
Manipal University, India
Chief Eng, Egypt
Modern Institute of Engineering & Technology, India
Cairo University, Egypt
Universiti Teknikal Malaysia Melaka, Malaysia
Universiti Malaysia Perlis, Malaysia
Riphah International University, Pakistan
Telangana University, India
Vellore Institute of Technology, India
Abdelmalek Essaadi University, Morocco
Egyptian Chinese University, Egypt
Loughborough University, UK

Bala Modi,	Gombe State University, Nigeria
Barhoumi Walid,	SIIVA-LIMITIC Laboratory, Tunisia
Basant Kumar Verma,	G H Raisonni College of Engineering, India
Benyettou Mohammed,	University center of Relizane, Algeria
Beshair Alsiddiq,	Prince Sultan University, Saudi Arabia
Bouchra Marzak,	Hassan II University, Morocco
Brahim Lejdel,	University of El-Oued, Algeria
Cagdas Hakan Aladag,	Hacettepe University, Turkey
Carlos Becker Westphall,	Federal University of Santa Catarina, Brazil
Chandrashekhar Bhat,	MIT Manipal, India
Changsoo Je,	Sogang University, South Korea
Chang-Yong Lee,	Kongju National University, South Korea
Cheng Siong Chin,	Newcastle University, Singapore
Cherkaoui Leghris,	Hassan II University of Casablanca, Morocco
Chikh Mohamed Amine,	University of Tlemcen, Algeria
Christian Mancas,	Ovidius University, Romania
Chuan-Ming Liu,	National Taipei University of Technology, Taiwan
Claude Tadonki,	MINES ParisTech-PSL, France
Dac-Nhuong Le,	Haiphong University, Vietnam
Dario Ferreira,	University of Beira Interior, Portugal
Dariusz Barbucha,	Gdynia Maritime University, Poland
Demian Antony D'Mello,	Canara Engineering College, India
Dimitris Kanellopoulos,	University of Patras, Greece
Dinesh Reddy,	SRM University, India
Domenico Rotondi,	Fincons SpA, Italy
Donatella Giuliani,	University of Bologna, Italy
Douglas Alexandre Gomes Vieira,	Enacom, Brazil
Edmund Lai,	Auckland University of Technology, New Zealand
El Murabet Amina,	Abdelmalek Essaadi University, Morocco
Elzbieta Macioszek,	Silesian University of Technology, Poland
Emir Kremic,	Federal Institute of Statistics, Bosnia and Herzegovina
F. Abbasi,	Islamic Azad University, Iran
Fabio Gasparetti,	Roma Tre University, Italy
Farshad Safaei,	Shahid Beheshti University, Iran
Felix J. Garcia Clemente,	University of Murcia, Spain
Fernando Tello Gamarra,	Federal University of Santa Maria, Brazil
Fernando Zacarias Flores,	Universidad Autonoma de Puebla, Mexico
Fiza Saher Faizan,	Dhacss Beachview Campus, Pakistan
Francesco Zirilli,	Sapienza Universita di Roma, Italy
Gajendra Sharma,	Kathmandu University, Nepal
Govindraj B Chittapur,	Basaveshwar Engineering College, India
Grigorios N. Beligiannis,	University of Patras, Greece
Grzegorz Sierpinski,	Silesian University of Technology, Poland
Gulden Kokturk,	Dokuz Eylul University, Turkey
Guzouli Larbi,	Higher National School of Renewable Energy, Algeria
H.Amca,	Eastern Mediterranean University, Turkey
Hacer Yalim Keles,	Ankara University, Turkey
Hadi Amirpour,	Universidade da Beira Interior, Portugal
Hala Abukhalaf,	Computer Application Technology, Palestine
Hamid Ali Abed AL-Asadi,	Iraq University College, Iraq
Hamid Khemissa,	USTHB University Algiers, Algeria
Hanming Fang,	Logistical Engineering University, China

Hao-En Chueh,	Yuanpei University, Taiwan
Hassan Ugail,	University of Bradford, UK
Hatem Yazbek,	Nova Southeastern University, USA
Hedayat Omidvar,	National Iranian Gas Company, Iran
Hicham Toumi,	University Hassan II Casablanca, Morocco
Hiromi Ban,	Sanjo City University, Japan
Huang Lianfen,	Xiamen University, China
Hyunsung Kim,	Kyungil University, Korea
Ilango Velchamy,	CMR Institute of Technology, India
Ilham Huseyinov,	Istanbul Aydin University, Turkey
Isa Maleki,	Science and Research Branch, Iran
Ismail Rakip Kara,	Karabuk University, Turkey
Israa Shaker Tawfic,	Ministry of Migration and Displaced, Iraq
Issa Atoum,	The World Islamic Sciences and Islamic Studies, Jordan
Issac Niwas Swamidoss,	Nanyang Technological University, Singapore
Iyad Alazzam,	Yarmouk University, Jordan
Jabbar,	Vardhaman College of Engineering, India
Janusz Kacprzyk,	Systems Research Institute, Poland
Jawad K. Ali,	SMIEEE, University of Technology, Iraq
Jesuk Ko,	Universidad Mayor de San Andres, Bolivia
Jia Ying Ou,	York University, Canada
Jing Zhang,	Harbin Engineering University, China
Jinguang Han,	Southeast University, China
Jiunn-Lin Wu,	National Chung Hsing University, Taiwan
Joan Lu,	University of Huddersfield, UK
Joao Antonio Aparecido Cardoso,	The Federal Institute of São Paulo, Brazil
Joao Calado,	Instituto Superior de Engenharia de Lisboa, Portugal
Jonah Lissner,	technion - israel institute of technology, Israel
Jong-Ha Lee,	Keimyung University, South Korea
José Manuel Fonseca,	NOVA University of Lisbon, Portugal
Jun Hu,	Harbin University of Science and Technology, China
Juntao Fei,	Hohai University, P. R. China
Kamel Benachenhou,	Blida University, Algeria
Kanstantsin MIATLIUK,	Bialystok University of Technology, Poland
Ke-Lin Du,	Concordia University, Canada
Keneilwe Zuva,	University of Botswana, Botswana
Kenjiro T. Miura,	Shizuoka University, Japan
Kevin Matthe Caramancion,	University at Albany, New York
Khader Mohammad,	Birzeit University, Palestine
khaled Osama Elzoghaly,	Alexandria University, Egypt
Khalid M.O Nahar,	Yarmouk University, Jordan
Kholladi Mohamed-Khireddine,	Lakhdar University of El Oued, Algeria
Kire Jakimoski, FON University,	Republic of Macedonia
Kirtikumar Patel,	Hargrove Engineers and Constructors, USA
Klenilmar L. Dias,	Federal Institute of Amapa, Brazil
Koh You Beng,	University of Malaya, Malaysia
Liquan Chen,	Southeast University, China
Loc Nguyen,	Loc Nguyen's Academic Network, Vietnam
Luisa Maria Arvide Cambra,	University of Almeria, Spain
M A Jabbar,	Vardhaman College of Engineering, Hyderabad
M V Ramana Murthy,	Osmania University, India
M. Akhil Jabbar,	Vardhaman College of Engineering, India

Mabroukah Amarif,	Sebha University, Libya
Mahdi Sabri,	Islamic Azad University, Iran
Malka N. Halgamuge,	The University of Melbourne, Australia
Mallikharjuna Rao K,	IIIT Naya Raipur, India
Mamoun Alazab,	Charles Darwin University, Australia
Marcin Paprzycki,	Polish Academy of Sciences, Poland
Marco Javier Suarez Baròn,	Pedagogical and technology of Colombia, Colombia
Mario Versaci,	DICEAM - Univ. Mediterranea, Reggio Calabria
Maumita Bhattacharya,	Charles Sturt University, Australia
Meenakshi Sharma,	Galgotias university, India
Mehdi Nezhadnaderi,	Islamic Azad University, Iran
Mesfin Abebe Haile,	Adama Science and Technology University, Ethiopia
Michail Kalogiannakis,	University of Crete, Greece
Mihai Carabas,	University POLITEHNICA of Bucharest, Romania
Mihai Horia Zaharia,	Gheorghe Asachi Technical University, Romania
Mirsaeid Hosseini Shirvani,	Islamic Azad University, Iran
Mohamed Ismail Roushdy,	Ain Shams University, Egypt
Mohammad A. Alodat,	Sur University College, Oman
Mohammad Ashraf Ottom,	Yarmouk University, Jordan
Mohammad Hamdan,	Heriot Watt University, UAE
Mohammad Jafarabad,	Iran University of Science & Technology, Iran
Mohammed Bouhorma,	Abdelmalek Essaadi University, Morocco
Mohammed Fatehy,	Soliman University of Urmia, Iran
Mohammed Mahmood Ali,	Osmania University, India
Morteza Alinia Ahandani,	University of Tabriz, Iran
Mourad Chabane Oussalah,	University of Nantes, France
Mu-Chun Su,	National Central University, Taiwan
Mueen Uddin,	Universiti Brunei Darussalam, Brunei Darussalam
Muhammad Arif,	Guangzhou University, China
Muhammad Sarfraz,	Kuwait University, Kuwait
Mu-Song Chen,	Da-Yeh University, Taiwan
Nabila labraoui,	University of TLEMCEM, Algeria
Nadia Abd-alsabour,	Cairo university, Egypt
Nahlah Shatnawi,	Yarmouk University, Jordan
Nalin D. K. Jayakody,	National Research Charles Sturt University, Australia
Natarajan Meghanathan,	Jackson State University, USA
Nikola Ivkovic,	University of Zagreb, Croatia
Nikolai Prokopyev,	Kazan Federal University, Russia
Nisheeth Joshi,	Banasthali University, India
Nivedita Bhirud,	VIIT, India
Olakanmi Oladayo O,	University of Ibadan, Nigeria
Oleksii K. Tyschenko,	University of Ostrava, Czech Republic
Osama Hosam,	Taibah University, Saudi Arabia
Otilia Manta,	Romanian American University (RAU), Romania
P Joseph Charles,	St. Joseph's College, India
P.V.Siva Kumar,	VNR VJIET, India
Paolo Dario,	Scuola Superiore Sant'Anna, Italy
Pasupuleti Venkata Siva Kumar,	VNR VJIET, India
Paulo Jorge dos Mártires Batista,	University of Évora, Portugal
Pavel Loskot,	Swansea University, UK
Pi-Chung Hsu,	Shu-Te University, Taiwan
Piotr Kulczycki,	Systems Research Institute, Poland

Pokkuluri Kiran Sree,	Sri Vishnu Engineering College for Women, India
Przemyslaw Falkowski-Gilski,	Gdansk University of Technology, Poland
Quang Hung Do,	University of Transport Technology, Vietnam
Radha Raman Chandan,	Banaras Hindu University, India
Rakesh Kumar,	Dr.Sakunthala Engineering College, India
Ramadan Elaiees,	University of Benghazi, Libya
Ramgopal Kashyap,	Amity University Chhattisgarh, India
Rami Raba,	Al Azhar University, Palestine
Ray-I Chang,	National Taiwan University, Taiwan
Rinku Datta Rakshit,	Asansol Engineering College, India
Ripal D Ranpara,	Atmiya University, India
Rodrigo Pérez Fernández,	Universidad Politécnica de Madrid, Spain
Rosalba Cuapa Canto,	Universidad Autónoma de Puebla, México
Ruijiang Li,	eBay, China
S.Sridhar,	Easwari Engineering College, India
S.Vijayarani,	Bharathiar University, India
Saad Al-Janabi,	Al- hikma college university, Iraq
Sabyasachi Pramanik,	Haldia Institute of Technology, India
Sahar Saoud,	Ibn Zohr University, Morocco
Sahil Verma,	Chandigarh University, India
Said Agoujil,	Moulay Ismail University, Morocco
Saiqa Aleem,	Zayed University, U.A.E
Samir Kumar Bandyopadhyay,	University of Calcutta, India
Sarra Nighaoui,	National Engineering School of Tunis, Tunisia
Sayali Kulkarni,	Cummins College of Engineering, India
Sebastian Fritsch,	IT and CS enthusiast, Germany
Sébastien Combéfis,	ECAM Brussels Engineering School, Belgium
Seema Verma,	Banasthali University, India
Shah Khalid Khan,	RMIT University, Australia
Shah Nazir,	University of Swabi, Pakistan
Shahid Ali,	AGI Education Ltd, New Zealand
Shahnorbanun Sahran,	Universiti Kebangsaan, Malaysia
Shahram Babaie,	Islamic Azad University, Iran
Shashikant Patil,	SVKMs NMIMS, India
Shoeib Faraj,	Technical And Vocational University, Iran
Siarry Patrick,	University Paris-Est Creteil, France
Siddhartha Bhattacharyya,	CHRIST (Deemed to be University), India
Sikandar Ali,	China University of Petroleum, China
Siva Kumar,	VNR VJIET, India
Solomiia Fedushko,	Lviv Polytechnic National University, Ukraine
Sreenivasa Rao Ijjada,	Gitam Deemed to be University, India
Srinivas Bachu,	MLR Institute of Technology and Management, India
Stefano Michieletto,	University of Padova, Italy
Subhendu Kumar Pani,	Krupajal Engineering College, India
Suhad Faisal Behadili,	University of Baghdad, Iraq
Sujatha,	Vellore Institute of Technology, India
Sukhdeep kaur,	punjab technical university, India
Sun-yuan Hsieh,	National Cheng Kung University, Taiwan
T V Rajini Kanth,	Professor & Dean R&D, India
Tanzila Saba,	Prince Sultan University, Saudi Arabia
Thenmalar S,	SRM Institute of Science and Technology, India
Titas De,	Data Scientist - Glance Inmobi, India

Usman Naseem,
V. Dinesh Reddy,
Varun Jasuja,
Veena M.N,
Viranjay M. Srivastava,
Walid Barhoumi,
Wenyuan Zhang,
William R. Simpson,
WU Yung Gi,
Xiao Wang,
Xiao-Zhi Gao,
Xinrong Hu,
Yacef Fouad,
Yamuna devi.N,
Yas A. Alsultanny,
Yousef Farhaoui,
Yousef J. Al-Houmaily,
Youssef Taher,
Yuan Tian,
Zeshui Xu,
Zeyar Aung,
Zhu Wang,
Zoran Bojkovic,

University of Sydney, Australia
SRM University, India
Guru Nanak Institute of Technology, India
P.E.S.College of Engineering, India
University of KwaZulu-Natal, South Africa
University of Carthage, Tunisia
Southeast University, China
Institute for Defense Analyses, USA
Chang Jung Christian University, Taiwan
Amazon, USA
University of Eastern Finland, Finland
Wuhan Textile University, China
Division Productique et Robotique, Algeria
Department of Computing, India
Uruk University, Iraq
Moulay Ismail University, Morocco
Institute of Public Administration, Saudi Arabia
Center of Guidance and Planning Rabat, Morocco
Nanjing Institute of Technology, China
Sichuan University, China
Khalifa University of Science and Technology, UAE
Hassan II University of Casablanca, Morocco
University of Belgrade, Serbia

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)



Artificial Intelligence Community (AIC)



Soft Computing Community (SCC)



Digital Signal & Image Processing Community (DSIPC)



10th International Conference on Signal, Image Processing and Pattern Recognition (SIPP 2022)

Applied Monocular Reconstruction of Parametric Faces with Domain Engineering.....	01-16
<i>Igor Borovikov, Karine Levonyan, Jon Rein, Pawel Wrotek and Nitish Victor</i>	
Cognitive Graphical Password based on Recognition with Improved User Functionality.....	17-24
<i>Mozhdeh Sarkhoshi and Qianmu Li</i>	
Expert Systems Generating Machine for Image Processing Applications.....	25-44
<i>Maan Ammar, Khuzama Ammar, Kinan Mansour and Waad Ammar</i>	
Phase Difference based Doppler Disambiguation Method for TDM-MIMOFMCW Radars.....	45-53
<i>Qingshan Shen and Qingbo Wang</i>	

3rd International Conference on Natural Language Processing & Computational Linguistics (NLPCL 2022)

Towards Modi Script Preservation: Tools for Digitization.....	55-67
<i>Kishor Patil, Neha Gupta, Damodar M and Ajai Kumar</i>	
Task-Oriented Dialogue Systems: Performance vs Quality-Optima, A Review..	69-87
<i>Ryan Fellows, Hisham Ihshaish, Steve Battle, Ciaran Haines, Peter Mayhew, J. Ignacio Deza</i>	
Emoji-based Fine-Grained Attention Network for Sentiment Analysis in the Microblog Comments.....	89-100
<i>Deng Yang, Liu Kejian, Yang Cheng, Feng Yuanyuan and Li Weihao</i>	

3rd International conference on Big Data, Machine learning and Applications (BIGML 2022)

A Big Data Driven System to Improve Residential Irrigation Efficiency using Machine Learning and AI.....	101-111
<i>Kai Segimoto, Nelly Segimoto and Yu Sun</i>	
FunReading: A Game-based Reading Animation Generation Framework to Engage Kids Reading using AI and Computer Graphics Techniques (for Special Needs).....	113-120
<i>Jiayi Zhang, Jiayu Zhang, Justin Wang and Yu Sun</i>	

AI in Telemedicine: An Appraisal on Deep Learning-based Approaches to Virtual Diagnostic Solutions (VDS).....229-243
Ozioma Collins Oguine and Kanyifeechukwu Jane Oguine

7th International Conference on Software Engineering (SOEN 2022)

A Real-time Multiplayer FPS Game using 3D Modeling and AI Machine Learning.....121-130
John Zhang and Yu Sun

10th International Conference on Artificial Intelligence, Soft Computing (AISC 2022)

Comparison of Forecasting Methods of House Electricity Consumption for Honda Smart Home.....131-140
Farshad Ahmadi Asl and Mehmet Bodur

A Social-Driven Intelligent System to Assist the Classification of Pet Emotions using Deep Learning and Big Data Analysis.....141-149
Hans Li and Yu Sun

7th International Conference on Networks, Communications, Wireless and Mobile Computing (NCWMC 2022)

Mass Surveillance, Behavioural Control, And Psychological Coercion the Moral Ethical Risks in Commercial Devices.....151-168
Yang Pachankis

12th International Conference on Computer Science and Information Technology (CCSIT 2022)

Integrating Ethical, Legal and Social Aspects into Common Procedure Models.....169-174
Sascha Alpers

Voice Chatbot for Hospitality.....175-184
Sagina Athikka and John Jenq

A Comparison between Vgg16 and Xception Models used as Encoders for Image Captioning.....185-195
Asrar Almogbil, Amjad Alghamdi, Arwa Alsahli, Jawaher Alotaibi, Razan Alajlan and Fadiah Alghamdi

Outlier Detection and Reconstruction of Lost Land Surface Temperature Data in Remote Sensing.....	197-205
<i>Muhammad Yasir Adnan, Yong Xue and Richard Self</i>	
A Method to Compactly Store Scrambled Data Alongside Standard Unscrambled Disc Images of CD-ROMs.....	207-215
<i>Jacob Hauenstein</i>	
An Overview of Phishing Victimization: Human Factors, Training and the Role of Emotions.....	217-228
<i>Mousa Jari</i>	

APPLIED MONOCULAR RECONSTRUCTION OF PARAMETRIC FACES WITH DOMAIN ENGINEERING

Igor Borovikov, Karine Levonyan, Jon Rein,
Pawel Wrotek and Nitish Victor

Electronic Arts, Redwood City, CA, USA

ABSTRACT

Many modern online 3D applications and videogames rely on parametric models of human faces for creating believable avatars. However, manual reproduction of someone's facial likeness with a parametric model is difficult and time-consuming. Machine Learning solution for that task is highly desirable but is also challenging. The paper proposes a novel approach to the so-called Face-to-Parameters problem (F2P for short), aiming to reconstruct a parametric face from a single image. The proposed method utilizes synthetic data, domain decomposition, and domain adaptation for addressing multifaceted challenges in solving the F2P. The open-sourced codebase illustrates our key observations and provides means for quantitative evaluation. The presented approach proves practical in an industrial application; it improves accuracy and allows for more efficient models training. The techniques have the potential to extend to other types of parametric models.

KEYWORDS

Face Reconstruction, Parametric Models, Domain Decomposition, Domain Adaptation.

1. INTRODUCTION

Modern virtual environments strive to deliver a life-like representation of human facial likeness under a limited computational budget. Such demand emerges both in the production art tools and user-facing applications. Customizable video games or 3D chat characters generated from user-provided photographs are in high demand (e.g., [13, 19, 23]). The following subsection explains some necessary terminology.

1.1. 3DMM vs. Parametric Model

The paper's context requires emphasizing some crucial distinctions in the avatar creation frameworks. Namely, there are two fundamentally distinct methods to creating human faces in Computer Graphics. One approach utilizes a fully Morphable 3D Mesh (3DMM) to adjust individual vertices to produce a desired shape [2]. A radically different approach is parametric (see [2,37]). A parametric model abstracts from the vertices and, in that sense, is more general. It relies on a pre-fixed collection of hand-authored construction elements (sometimes called "blendshapes"). These elements are used across all the characters within an application. Such elements contribute to the target shape with the weights defined by the input parameters, e.g., the distance between eyes, size of the mouth or nose, and alike.

A direct comparison of 3DMM vs. parametric frameworks for ML applications is rarely insightful within an industrial context: the engineering, production, and art style decisions take precedence in designing a concrete application. On the one hand, models can be hand-crafted with highly detailed 3DMMs in the film industry. On the other end, expensive authoring, tight development timeframe, limited memory, or bandwidth get better use with a modest set of predefined assets. Such assets can be preloaded to the customer hardware and allow for compact encoding of the models via a relatively small parameter set in interactive environments.

The paper focuses on *parametric* models suitable for interactive user-facing software and leaves out 3DMM-based systems for a different exploration.

1.2. Face-to-Parameters with a Heterogeneous Target Space

In our formulation, the Face-to-Parameters (F2P) problem aims to reproduce the face on a single input image in the best manner possible by optimally selecting two types of model parameters: continuous and discrete. The presence of discrete construction elements makes our target space heterogeneous.

Continuous parameters, as we already mentioned, are the weights of blendshapes. Blendshapes are a standard technique in modeling complex articulated objects, like human bodies and faces (blendshape methods reviewed in [39]). An application or a modeling tool has a fixed set of blendshapes used across all the characters. Given design limitations, the blendshapes aim to represent all anticipated target geometries in the best manner possible. All blendshapes contribute to the resulting geometry simultaneously and are not mutually exclusive. Some examples of a blendshapes effect on the output are the length of the nose, the shape of the nose (curvy, straight, up, down, wide, narrow), mouth location relative to the chin and the nose, the shape of the mouth (plum, thin, curved up or down), how prominent is the chin, and alike.

The discrete elements are fixed too but mutually exclusive within their region of the target geometry. They represent gross deformations of an underlying mesh. Possible examples are a distinct nose shape not reachable with the available blendshapes (say, a broken nose), a particular texture layer (e.g., an artist painted enhanced details for the features too fine to represent with the mesh), or both (e.g., a carnival mask).

In the ML context, fitting the continuous parameters is a regression problem and fitting the discrete elements is a classification problem. In this paper, the F2P problem combines both. In our proprietary application, the dimensionality of the continuous parameters space is ~ 100 , and the discrete elements count exceeds 300. The parameters by design come partitioned into several facial regions (e.g., nose, mouth, and alike), leading to the combinatorial complexity in the order of 10^{11} . For comparison, the FLAME model [12] has no discrete elements and exposes ~ 300 continuous parameters, with ~ 100 devoted to articulating the character. We exclude facial expressions from consideration in this paper. An open-source software Makehuman [15], which we use for reproducible quantitative evaluation of this paper, exposes over 100 continuous parameters and only a handful of discrete ones. Architecturally, Makehuman is the closest counterpart of the proprietary software we used for this work. That dictates our choice to open-source our experimental code using Makehuman rather than other systems like FLAME.

For the brevity of this presentation, we omit color palette elements, hairstyles, facial hair, makeup, tattoo, accessories, etc. - anything that does not directly change the geometry of a character's face. Also, we exclude ears and neck as they are frequently occluded in the imagery.

At the high level, the F2P problem formulation here is similar to that one in [23].

1.3. The Approach Outline

For the reasons we explain in the following sections, we cast the heterogeneous F2P problem as a classic supervised learning problem. That requires abundant training data, i.e., facial images with the corresponding parameters. Identifying the source of the training data requires careful navigation around privacy and licensing aspects. Most of the available human face datasets, e.g., a popular CelebA dataset [14], exclude commercial applications. With growing concerns around privacy, many previously accessible facial datasets are no longer available. The most direct way to work around these challenges is to use synthetic data. The generation of synthetic data is often straightforward and can produce practically unlimited amounts of it reasonably quickly with a wealth of the associated metadata.

The synthetic nature of the data creates new opportunities not readily available when working with real-world data. The parameters of the human face model naturally map to different facial regions. In parallel to that grouping, the rendering pipeline can include automated generation of the corresponding semantic segmentation of the synthetic images. Such natural separation leads to domain decomposition: the ML pipeline can become a hierarchical ensemble of models dealing with the general structure of the face and local models that control separate regions. The ensemble allows for smaller models that are less prone to overfitting, can train and execute in parallel utilizing data and model parallelism (e.g., [31] reviews conceptually similar works).

While synthetic data facilitates the decomposition and training of the related ML models, it also presents a challenge due to the inevitable domain gap between synthetic and real-world imagery (a brief introduction to domain adaptation in [40]). The imagery such as selfies may be only one of the possible input types to the F2P models. Examples of other possible target domains include sketches, fine art portraits, faces from comics, anime, and more. Previously unseen domains may degrade a naive F2P ML model's accuracy or render it unusable. That makes the domain gap issue broad and even more critical, suggesting that domain adaptation must be an integral part of the system we build. At the same time, the domain adaptation functionality must be modular and easily replaceable to accommodate different future applications.

We utilize style transfer for domain adaptation (see the seminal paper [4]). The direction of the style transfer is from the target imagery (e.g., selfies) to the synthetic images, which could have a distinct (fixed) art style. It is "inverse" in that sense. We train GANs specific to each of the multiple target domains and use them as an adapter. Training such GANs is independent of the decomposition-based ensemble training. We find that the decomposition and domain adaptation enhance each other and lead to an ensemble that produces better results in solving the F2P problem than a direct approach based on a single monolithic model.

1.4. Contribution

The paper emphasizes practical aspects of converting facial images into the parameters of a parametric model of a human face and proposes a novel, efficient solution. While the individual building blocks of our approach are not new (hierarchical decomposition, domain adaptation, models ensemble), their proposed combination is not found in the literature.

Concretely, we cast the F2P problem as a supervised learning problem on synthetic data, introduce a model ensemble to take advantage of the hierarchical nature of the domain, and train separate dedicated models for domain adaptation. The proposed architecture is different from the previous works (e.g., [23]) and offers several practical advantages in the industrial environment. The open-sourced code illustrates the claim that the proposed ensemble performs better than a

monolithic model thanks to leveraging the structure of the application domain. We advocate the modularity of the proposed system and offer a quantitative evaluation of claims.

In the rest of the paper, we review some of the previous works and then focus on the main objective: building an efficient ML system solving the outlined F2P problem, which infers the target face parameters from the input images. The system we develop is implemented as proprietary industrial software. However, we address reproducibility and quantitative evaluation of the techniques with Makehuman (free open-source software) by open-sourcing our experimental code as well [38].

2. PREVIOUS WORKS

Facial likeness reconstruction in computer vision is a vast active area of research. For a relatively recent comprehensive review of the field, we refer to [35]. Here, we highlight only several of the relevant publications. A well-established body of work aims at generating sculptable (and texturable) mesh directly from monocular input: e.g., [18,24,28,19,29,11,22] or video [26,27]. Parametric models may also utilize morphable meshes with a fixed topology where the parametric space explicitly encodes deltas for vertices subsets. Their reconstruction from 2D imagery achieves a very high accuracy [20] but conceptually is similar to 3DMM reconstruction. Such models differ from more specialized ones that use higher-level construction elements, e.g., blendshapes and bone-driven morphs, which are usually specific to a particular graphics engine. The models relying on high-level construction elements are somewhat less common in the literature even though they date back to the early days of Computer Graphics and Computer Vision, e.g., [16]. They provide a critical advantage in data compression, which motivated one of the early works [3]. Compact data representation remains in demand for interactive graphics applications as they strive for computational efficiency. Also, as the result of compact representation, parametric models allow for easier manual authoring compared to freely deformable meshes.

Parametric models heavily depend on the underlying character modeling system's proprietary nature. That creates an obstacle to transferring the parametric techniques across applications and renders a comparison to the 3DMM-based methods irrelevant. The additional difficulty comes from discrete elements of facial reconstruction. In our work, we must handle both continuous and discrete elements. The paper [23] uses a custom-trained differentiable renderer (DR), a powerful tool in the continuous domain. However, a DR could be more problematic for fitting discrete models and is still too heavy for mobile devices.

The paper [23] works around the domain gap between real and synthetic faces by introducing discriminative loss – a perceptual distance computed with embeddings of a facial recognition CNN. The discriminative part becomes an integral part of a larger model. For our purposes of building an ensemble with pluggable components, we explored several other more general GAN-based techniques. Our brief overview starts with [32] discussing various ways to perform domain adaptation. The Neural Style Transfer [4] is the pioneering work utilizing both content and style losses. The approach uses only a single image to represent the style. That appears to be rather limiting for our application. An early generative model applied for image-to-image translation between domains is pix2pix [7], and it uses conditional GANs. The conditional GAN introduces two extra losses [8] that address radical differences for the domains in question. However, it requires paired image-to-image translation. CycleGAN [34] works for unpaired data. It translates the source image into the target without paired examples by introducing the (computationally intensive) consistency loss. More recent advancement on style transfer is well-established StyleGAN2 [30] and its later modification StyleGAN3 [36]. The StyleGAN2 architecture has multiple advantages: the scale-based hierarchy for the generator and the discriminator, pretrained

models, and better stability than many other style transfer approaches. Its hierarchical nature helps to control the enhancement of the appearance of stylized images via the so-called layer-swapping. It allows for balancing features at different levels of details using blending weights, making it an attractive candidate for our exploration.

In conclusion of this section, we note that [23] has similar objectives to ours and some superficial similarities of techniques. However, the constraints and the methodology proposed here are quite different. Our target platform excludes a differentiable renderer from the potential toolset. We exclude pixel loss from consideration since we allow various projections and lighting conditions for the input images. Finally, the proprietary codebase of [23] makes reproduction and a direct comparison with its results difficult. As such, the cited work serves more as a valuable inspiration rather than a technical reference.

3. F2P AS A SUPERVISED ML PROBLEM

In this paper, we treat F2P as a classic supervised ML problem. The training and validation data are synthetic, generated by the target software. That eliminates any licensing or privacy concerns. The open-source Makehuman [15] provides similar functionality to our proprietary software, and we use it for the quantitative evaluation of this study with the codebase [38].

A simple way to generate the training data is to do it offline. An instrumented client application accepts a “recipe” (a complete set of parameters, including those not considered as target variables) for constructing a character as an input. Next, it renders and saves several views of the generated character. The views may vary by camera, lighting, pose, facial expression, and gaze direction. Here, we only limit the target space to the facial parameters and exclude the view and pose parameters. The normalized continuous parameters (blendshape weights) map to the floating-point target vector. The discrete components use one-hot encoding occupying target vector slices of the size corresponding to the number of the options.

A direct approach to training a model mapping an image to the heterogeneous target is to train a CNN with multi-part loss function L :

$$L = \sum_{i=1}^N v_i R_i + \sum_{i=1}^N w_i C_i \quad (1)$$

Here, N is the number of facial regions (seven in our main case study and we keep only three for Makehuman experimentation). R_i is either mean L2 or L1 loss for modifiers, C_i is the cross-entropy loss for the discrete elements, and v_i , and w_i weights that can be adaptive [5]. We use transfer learning with *inception3* [25] and *squeezenet* [6,33] from Pytorch [17] models zoo as the starting point. Transfer learning provides a reasonable accuracy relatively early in the training process, but complete training may take longer with ~10k input images randomized across all target dimensions. The relatively long training times in the direct approach are a disadvantage slowing iterations on the models for the evolving customer product. Also, a monolithic model makes it hard to address inaccuracies in concrete facial regions. Finally, the amount of training data may be insufficient for reliable generalization due to the high dimensionality of the target space, potentially leading to various biases and overfitting. These considerations motivated our decomposition approach.

4. ENSEMBLE ENGINEERING VIA DOMAIN DECOMPOSITION

The natural subdivision of a human face into regions like the nose, mouth, and alike leads to the target space's corresponding decomposition. Such decomposition is present in many parametric-based modeling applications, including Makehuman. Equation (1), grouped by region, gives loss functions for local models with the following caveats.

The first caveat is overcompleteness, i.e., a parametric model may generate the same visual appearance with different parameter values. One obvious source of such over-completeness is scale. Fixing a particular scale variable in the training data normalizes it and eliminates over-completeness in our studies. Next, a group loss corresponding to a specific facial region may include local (e.g., angle) and global features (e.g., placement of the feature relative to the other facial features). Manual engineering of the group loss to account for such subdivision could be error-prone and subject to frequent revisions as the client software evolves. We address local-vs-global ambiguity by introducing weights into the ensemble as follows. Global features learned within local groups would result in predictions equal to an average value over the dataset. When learned as part of the aggregate complete model, they will result in a much better prediction. Their learned combination with the introduced weights will automatically reflect their roles. One way to formulate training for the ensemble weights is to frame it as an optimization problem

$$E(w) = \left| \sum_{i=1}^{n_k} w_i^{(k)} M_i^{(k)}(D) - T(D) \right|^2 \rightarrow \min_w \quad (2)$$

Here, $E(w)$ is the cumulative L2 error computed for weights $w^{(k)}_i$ corresponding to the group k and target variable i . By running models $M^{(k)}$ on the training dataset D , we obtain predictions for each group k . Linear combination of the predictions with weights $w^{(k)}_i$ gives ensemble prediction. We compare it with the known target vector $T(D)$ and compute mismatch on the entire dataset as a function of w . The weights vectors $w^{(k)}_i$ dimensionality equals the dimensionality of the target parameters space. Also, we normalize the weights so that they sum to 1 for each target variable. That gives complementary weights w_i and $1-w_i$ for the coordinates shared by local and global features. Solving (2) for w is a straightforward coordinate-wise task. Not restricting weights to positive values allows to utilize consistent bias in either local or global models and can further improve the accuracy of the ensemble. Figure 1 summarizes our approach to constructing, training, and using the decomposition-based ensemble for inference.

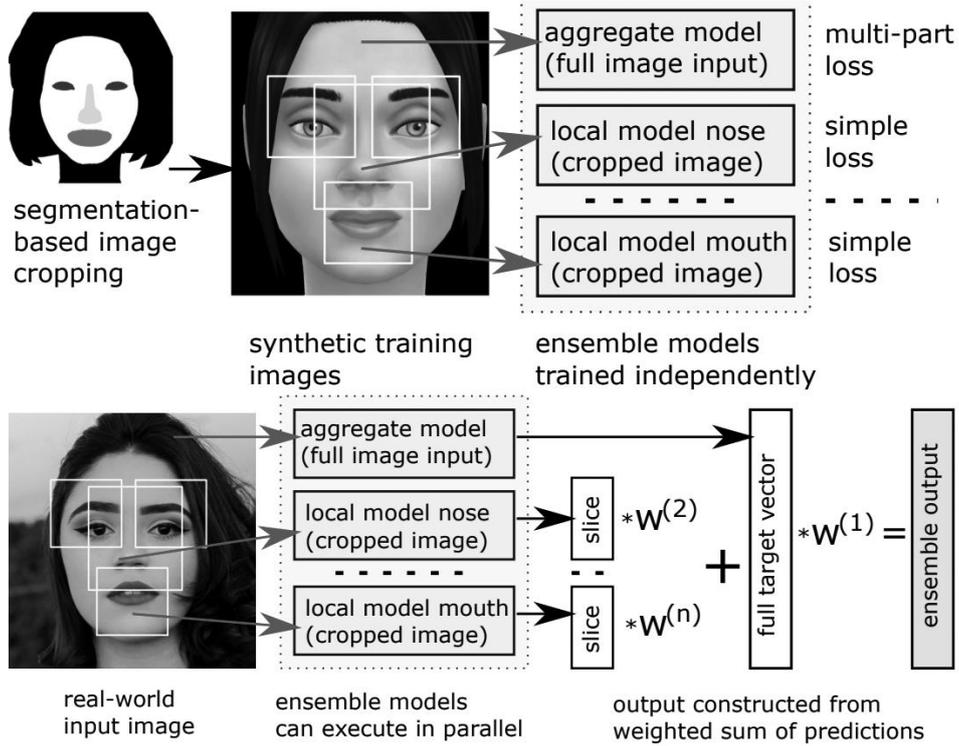


Figure 1. Training (top) and inference (bottom) with the decomposition-based ensemble. (We defer domain adaptation part until the upcoming section, Figure 5.)

A complete target space in our study has dimensionality close to 400. Its combinatorial part has a complexity of $\sim 10^{11}$. Decomposition into subspaces radically reduces the dimensionality of the target spaces for each model we need to train. In our application, the largest modifiers subspace has dimensionality ~ 20 . The largest number of discrete options for discrete elements for a region is under 100. Since the subspaces for discrete parts represent mutually exclusive choices, the combinatorial complexity reduces by many orders of magnitude. Combined with lower input resolution for local features, these factors allow for more efficient training with a smaller amount of training data necessary for sufficient sampling coverage of the target space. Since training models for the decomposed local features are less demanding than for an aggregate model, we may use smaller CNNs. In our experiments, we use the Pytorch model zoo's *squeezenet* for local features inference. Its smaller input resolution, 224×224 , is suitable for cropped input image parts. Also, *squeezenet* is a lightweight model at only $\sim 3\text{Mb}$ vs. *inception3*, which is close to 100Mb and takes 299×299 inputs. Our main study has 13 models, two per region (regression and classification grouped separately), with one region lacking discrete elements.

5. ENSEMBLE EVALUATION WITH AN IN-HOUSE SOFTWARE

In our first round of experimentation, the proprietary software has a stylized, cartoonish rendering, defined by the art direction. Using automated objective evaluation (e.g., cosine distance in the latent space of FaceNet [21]) while disregarding the artistic style is problematic due to the domain gap. All our training and evaluation datasets are synthetic, rendered in the same artistic style, while we expect that input images would come from photographs. To workaround, we use a panel-of-experts method. Nine experts evaluate the reconstruction quality of a hand-picked set of unannotated 20 facial images representing various ages, ethnicities, lighting, image quality, poses, and facial expressions. During the evaluation, the respondents,

besides other things, have to rate each of the seven major facial regions of the reconstructed characters with binary good/bad evaluation (could be less informative than the Likert scale but makes the questionnaire easier and faster to complete).

Table 1 summarizes the nine experts' responses in both rounds of the assessment: one for the naive aggregate model and the other for the models' ensemble. The panel includes both professional artists and technical staff members. The table highlights improvements in the selected quality metrics obtained by introducing the model ensemble and clearly shows the advantages gained from the decomposition approach.

Table 1. Comparison of ensemble vs. aggregate models by a panel of experts. Models' decomposition and ensembling reduce reconstruction defects by 30-50%. The worst possible defects score is 180 for nine experts and 20 images. The FaceNet cosine distance to the input image also reduces on most images by ~10%.

Reconstruction defects by region	Cheeks	Chin	Eyes	Forehead	Jaw	Mouth	Nose
Naive aggregate model	19	18	69	3	29	37	34
Ensemble with models' decomposition	6	11	34	1	16	24	17



Figure 2. Naive aggregate model reconstruction (second on the left column) compared to the decomposition with models ensemble (second on the right). The numbers indicate FaceNet [21] cosine distance between the corresponding pairs. The imagery does not intend to represent any current or future commercial product. Photography attribution (side columns) [1].

6. DOMAIN GAP AND DOMAIN ADAPTATION

Naturally, we cannot ensure that synthetic training imagery covers the entire domain of the expected inputs or is its representative sub-domain. Also, in applications, we may encounter new domains not anticipated during training. The domain gap between synthetic and target imagery makes the models underperform.

Two major factors are leading to the domain gap. One is the limitations of the parametric model itself, which may not be powerful enough to generate a sufficient variety of faces. The second factor is the artistic stylization produced by the client application. The stylization lends itself to various domain adaptation techniques. Using style adaptation (e.g., see [4]) as an intermediate step between the input and the rest of the inference pipeline should reduce accuracy loss. Namely, we propose using an "inverse" style adaptation from the input to stylized synthetic imagery. We can train such an adapter for various inputs besides real-world images, e.g., fine art, anime, sketches.

After assessing various style transfer options (see Previous Works section), we use StyleGAN2 [30] for the academic part of our study (it has no commercial license) to train and evaluate the proposed style adapter in application to our F2P inference pipeline. The synthetic training domain in our experiments comes from Makehuman [15], picked for its architectural similarity with our proprietary software. The target domain is real-world imagery. For inference, we apply StyleGAN2 to the input image to make it look like it comes from Makehuman, then feed it into the pipeline trained with Makehuman synthetic dataset. To train the adapter, we start with a StyleGAN2 pretrained on the FFHQ dataset (created for [9]) and fine-tune it on 4000 Makehuman-generated images. We normalize the images by registering (resizing and aligning) them using dlib landmarks [10] to match the images' alignment in the FFHQ dataset. After computing the style weights, the inference continues as follows. We start with a normalized real image. Next, we compute the latent projection vector for the given image through the StyleGAN2 mapping network and then apply the blended style weights from the selected resolution to map the real image to the stylized image. Figure 3 illustrates the results of that process. Finally, we feed the stylized image to the inference F2P pipeline, which becomes an ensemble including the domain adaptation step.



Figure 3. Stylized images from unsplash.com to the proprietary and Makehuman styles. Top row: Makehuman style, middle: original photos, Bottom row: proprietary style.

7. QUANTITATIVE METRICS AND ABLATION STUDY

We conduct a series of experiments with Makehuman [15] software, reproducing the setup of our proprietary-based experimentation to support our findings using open-sourced, reproducible, quantifiable methods. For simplicity, we use a trivial baseline for the collected metrics: a mean over the evaluation dataset for the target variables. In our case, it corresponds to zero for all target variables describing an average face shape. Here, we focus on the regression part of the problem since Makehuman does not offer many discrete elements for the classification part. Besides, the decomposition advantage for classifiers follows directly from the combinatorial considerations.

Note that our goal for this concrete open-source study is not to train the best CNNs possible. That requires more effort and experimentation. Instead, we fix the training setup (meta-parameters, adaptive multi-part loss tuning, and the training schedule) to a reasonable common one and compare results in terms of accuracy between the combinations of the binary factors we describe after stating the following two claims to verify:

- **A decomposition-based ensemble** improves inference accuracy over the monolithic model.
- **Style adaptation via inverse style transfer** as a preprocessing step improves the accuracy of a model or models trained with synthetic training data.

For the decomposition-based ensemble claim, the goal is to obtain metrics characterizing the accuracy of the models trained either as a monolithic or decomposition-based ensemble. We evaluate the models utilizing the pretrained *squeezenet*, which we update in feature extraction and fine-tuning transfer learning modes. The motivation for considering both comes shortly. Overall, these three binary factors influence the results:

1. **Target space partitioning:** a complete target vector or limited to a particular feature group (e.g., nose parameters only). Such partitioning corresponds to the grouping terms of the target variables by features in multi-part loss (1).
2. **Input image:** either a complete frame or the corresponding to a feature group “semantic” crop. We resize the cropped input image to match the CNN input.
3. **Transfer learning mode:** feature extraction or fine-tuning. In our tests, the fine-tuning phase starts from the CNN obtained by the features extraction step to speed up training.

Figure 4 illustrates the setup for binary factors we test, and Table 2 summarizes the results.

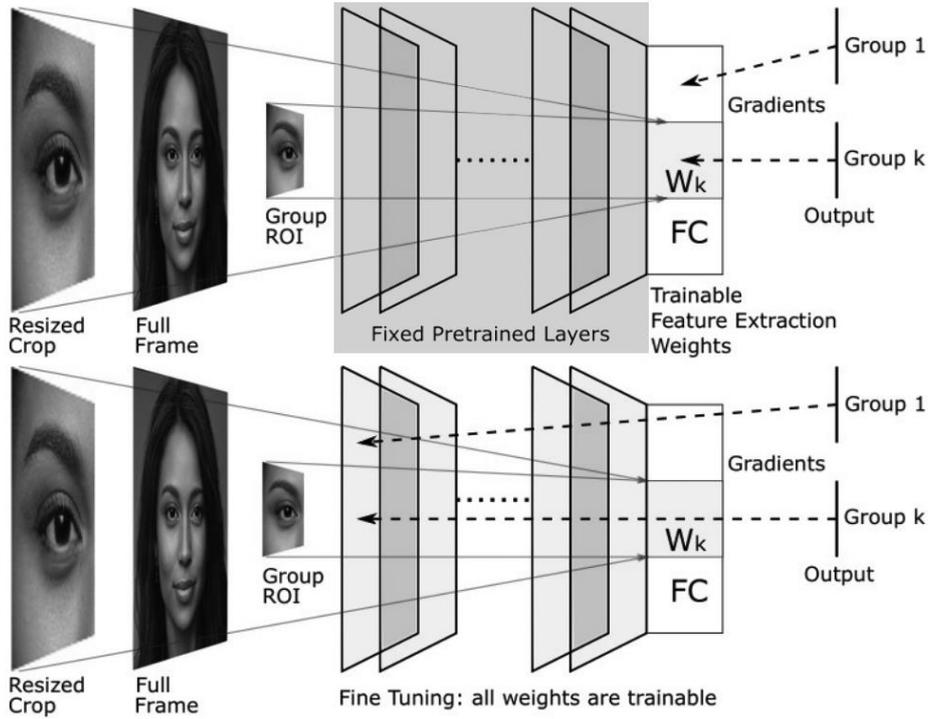


Figure 4. Quantitative evaluation setups and ablation study. We compare pairs: full-frame vs. crop input and complete multi-part loss vs. local group loss. Top: transfer learning with feature extraction. Bottom: transfer learning with fine-tuning.

We start with the target space partitioning. One of the factors that may influence the models' training and the resulting accuracy is the difference in scale of the gradients in the multi-part loss (1). If such a difference is significant, it may take longer to achieve the required accuracy across all target variables. We train the first set of models using feature extraction transfer learning to isolate that factor. In that form, the CNN's hidden layers' weights remain fixed with only the top fully-connected (FC) layer trained to fit the dataset. The input in this experiment is a full-frame image for all models.

When training stops with a predefined learning rate schedule by reaching a plateau for the evaluation dataset, we expect similar results between the local and monolithic models compared to the local target features. Running training sufficiently long with a shared learning rate scheduler should reduce the different gradient scales' effect (regardless of the overfitting concerns). The top part of Figure 4 illustrates the proposed setup. The hidden layers inside the greyed box remain fixed. The grey FC weights W_k for the target group k train (nearly) independently in monolithic and local features variants. The main expected difference is the loss improvement rate but similar resulting accuracy. Hence, the decomposition-based ensemble should not provide a notable advantage over the monolithic model with feature extraction training with a full-frame input. Table 2 shows only a tiny improvement from complete to partial loss function when using the full-frame as an input, supporting the intuition.

Table 2. Local feature group models demonstrate the advantages of the hierarchical decomposition of the F2P problem. The inaccuracy (mean L1- loss) generally decreases left-right top-down within each group.

The inaccuracy shown is relative to the baseline. Limiting the number of terms in multi-part loss and cropping the feature improves the resulting model by order of magnitude for some features. The resulting models' ensemble is far superior to the aggregate model trained on the complete input image while sharing with the aggregate one a similar training setup.

Features Group	Loss	Input	Inaccuracy vs. baseline (smaller is better)	
			Feature Extraction	Fine Tuning
Nose	Complete	Full frame	0.0001	-0.0039
	Local	Full frame	-0.0005	0.0007
		Cropped	-0.0059	-0.0743
Mouth	Complete	Full frame	-0.0001	-0.0135
	Local	Full frame	0.0005	0.0000
		Cropped	-0.0080	-0.0744
Eyes	Complete	Full frame	-0.0003	-0.0235
	Local	Full frame	-0.0001	0.0010
		Cropped	-0.0061	-0.0530

The other factor influencing the accuracy of the models is choosing the input image used to train the local feature CNN. We expect the results to improve by moving from a complete full frame to crops specific to the particular feature groups. That should work even in the feature-extraction transfer learning case. The Feature Extraction column in Table 2 confirms such an expectation. E.g., for the Nose features group, compare numbers in the feature extraction column between Full-Frame and Crop rows.

The progression from complete to local loss while using full-frame input does not improve accuracy much compared to the local loss and cropped images. Moving from feature extraction to fine-tuning, we adjust all weights in the CNN; see the bottom row of Figure 4. That makes learning the target features less constrained and provides a significant boost to the models' accuracy trained with local loss function corresponding to the feature image crop.

Moving from the full-frame to the cropped image input provides the most improvement for fine-tuning. The bold numbers in Table 2 correspond to the proposed local models used for constructing the decomposition-based ensemble. They show the best accuracy across the evaluated combinations of the binary factors.

Table 3. Fitting weights for the features individually reduces the prediction error compared to the simple ensemble with constant weights across all the dimensions. We set the baseline for this table per training type as the corresponding accuracy (validation loss) computed with constant weights (i.e., 0.0, 0.5, and 1.0). Smaller is better.

Constant weights	Feature Extraction		Fine Tuning	
	Full Frame	Crop	Full Frame	Crop
0.0 (aggregate model only)	-0.08	-0.09	-0.11	-0.12
0.5 (equal mix of aggregate and local models)	-0.08	-0.11	-0.23	-0.27
1.0 (local models only)	-0.08	-0.19	-0.44	-0.52

Finally, Table 3 shows that an ensemble with learned weights for its individual parameters outperforms its components or a weighted sum with fixed equal weights for the predictions.

We conclude this section with a discussion of domain adaptation. The inverse style transfer preprocessing step with StyleGAN2 trained on Makehuman images improves the ensemble's

accuracy. We could not utilize validation loss for that evaluation. Instead, we use cosine distance on FaceNet embeddings [20]. The absolute distances shown in Table 4 may look substandard than the commonly accepted threshold of 0.51 for person identification. However, their values are meaningful only in relative terms and illustrate that the domain adaptation moves the input distribution in the right direction, beneficial for the model ensemble. The bottom row distances for full-frame and crop cases are smaller for the style-adjusted images than for the original ones.

Table 4. FaceNet cosine distance to the original input image shows the advantages of the style transfer as the domain adaptation step. Here we use a subset of CelebA [14].

FaceNet cosine distance	Input	Feature Extraction		Fine Tuning	
		Full Frame	Crop	Full Frame	Crop
Original	[0.0]	0.889 \pm 0.13	0.893 \pm 0.12	0.917 \pm 0.12	0.875 \pm 0.12
Stylized	0.495 \pm 0.11	0.889 \pm 0.10	0.895 \pm 0.11	0.892 \pm 0.11	0.868 \pm 0.12

The presented experiments support our claims that in addition to being more manageable, easier to train, the ensemble-based approach takes advantage of the underlying properties of the models we use. It helps to train a better accuracy inference pipeline more practically.

8. DISCUSSION AND FUTURE WORK

For completeness, we mention here the limitations of the proposed technique. Reconstruction of a human face in the F2P problem is a multi-faceted task. The paper covers only a single subject - decomposition of "geometric" features that include continuous and discrete elements. In our experience, the classification of glasses, earrings, and some other localized features also benefit from the segmentation and decomposition approach. However, many "spread" features (e.g., hairstyle) do not easily lend themselves to the proposed method. Also, an assumption of the target features' independence is an oversimplification. The human faces exhibit strong correlations between age, gender, and ethnicity features. Including such correlations in the proposed ensemble may further improve the accuracy. One way of doing it is using a Bayesian framework over local models. It can replace a simple weighted sum with a belief propagation network and integrate otherwise ignored correlations between local features. It appears to be a valid subject to explore next. Other parametric models besides human faces may benefit from the proposed methodology.

9. CONCLUSION

The paper proposes a novel combination of well-established domain manipulation techniques summarized on Figure 5. Despite its conceptual simplicity, the domain decomposition combined with domain adaptation provides several measurable benefits in the F2P problem that are particularly valuable in the applications to both external user-facing software and the in-house art production pipelines. It facilitates training of the models, offers better control over their accuracy, convenient maintenance vital for industrial applications, smaller memory footprint during inference, and flexibility across input domains. The proposed approach is not fundamentally limited to parametric faces and may work for similar problems in other computer vision applications.

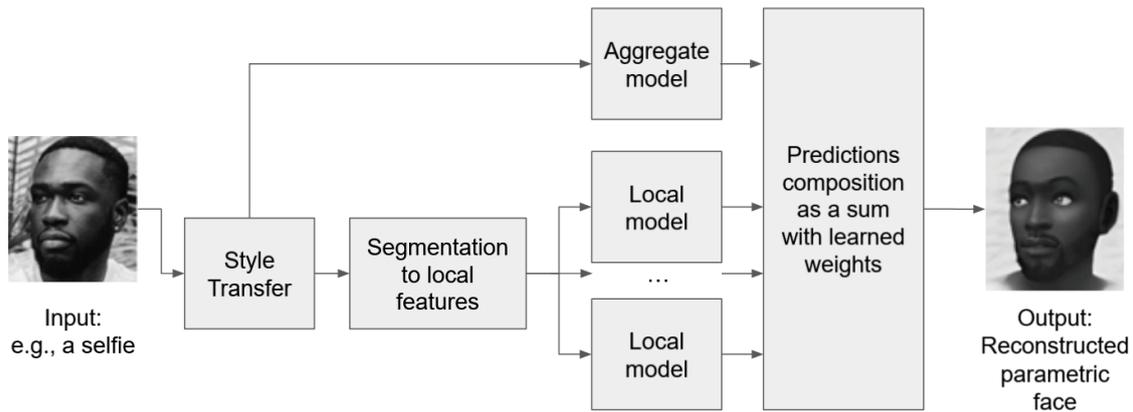


Figure 5. The diagram summarizes the proposed inference pipeline based on domain adaptation and decomposition ensemble. Style transfer, like the local models, is a pluggable module with the corresponding model easily re-trainable as the requirements change.

REFERENCES

- [1] Unsplash.com stock images with an all-permissive non-competing license., <https://unsplash.com2>.
- [2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 1999
- [3] Ahlberg, J.: Extraction and coding of face model parameters (1999)
- [4] Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. CoRRabs/1508.06576(2015), <http://arxiv.org/abs/1508.06576>
- [5] Heydari, A.A., Thompson, C., Mehmood, A.: Softadapt: Techniques for adaptive loss weighting of neural networks with multi-part loss functions. ArXivabs/1912.12355(2019)
- [6] Iandola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 1mb model size. ArXivabs/1602.07360(2017)
- [7] Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CoRRabs/1611.07004(2016), <http://arxiv.org/abs/1611.07004>
- [8] Kang, S., Ok, Y., Kim, H., Hahn, T.: Image-to-image translation method for game-character face generation. In: 2020 IEEE Conference on Games (CoG). pp. 628–631(2020). <https://doi.org/10.1109/CoG47356.2020.9231650>
- [9] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 4396–4405 (2019)
- [10] King, D.: Dlib-ml: A machine learning toolkit. J. Mach. Learn. Res.10, 1755–1758(2009)
- [11] Lattas, A., Moschoglou, S., Gecer, B., Ploumpis, S., Triantafyllou, V., Ghosh, A., Zafeiriou, S.: AvatarMe: Realistically Renderable 3D Facial Reconstruction “In-the-Wild. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 757–766 (2020)
- [12] Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, (Proc. SIGGRAPHAsia)36(6), 194:1–194:17 (2017), <https://doi.org/10.1145/3130800.3130813>
- [13] Lin, J., Yuan, Y., Zou, Z.: Meingame: Create a game character face from a single portrait. ArXiv abs/2102.02371 (2021)
- [14] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (December 2015)
- [15] <http://www.makehumancommunity.org> (2001-2022)
- [16] Parke, F.: A Parametric Model of Human Faces. Ph.D. thesis, University of Utah, Salt Lake City (1974)
- [17] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)

- [18] Richardson, E., Sela, M., Or-El, R., Kimmel, R.: Learning detailed face reconstruction from a single image. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 5553–5562 (2017)
- [19] Saito, S., Simon, T., Saragih, J., Joo, H.: Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 81–90 (2020)
- [20] Sanyal, S., Bolkart, T., Feng, H., Black, M.: Learning to regress 3D face shape and expression from an image without 3D supervision. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 7763–7772 (Jun 2019)
- [21] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 815–823 (2015)
- [22] Shang, J., Shen, T., Li, S., Zhou, L., Zhen, M., Fang, T., Quan, L.: Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry consistency. In: ECCV (2020)
- [23] Shi, T., Yuan, Y., Fan, C., Zou, Z., Shi, Z., Liu, Y.: Face-to-parameter translation for game character auto-creation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 161–170 (2019)
- [24] Song, M., Tao, D., Huang, X., Chen, C., Bu, J.: Three-dimensional face reconstruction from a single image by a coupled RBF network. *IEEE Transactions on ImageProcessing*21(5), 2887–2897 (2012). <https://doi.org/10.1109/TIP.2012.2183882>
- [25] Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2015 IEEE conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–9 (2015). <https://doi.org/10.1109/CVPR.2015.7298594>
- [26] Tewari, A., Bernard, F., Garrido, P., Bharaj, G., Elgharib, M.A., Seidel, H., Pérez, P., Zollhöfer, M., Theobalt, C.: Fml: Face model learning from videos. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 10804–10814 (2019)
- [27] Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M.: Face2face: real-time face capture and reenactment of RGB videos. *ArXiv abs/2007.14808*(2019)
- [28] Tran, A., Hassner, T., Masi, I., Medioni, G.: Regressing robust and discriminative 3d morphable models with a very deep neural network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1493–1502 (2017)
- [29] Tran, L., Liu, F., Liu, X.: Towards high-fidelity nonlinear 3d face morphable model. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1126–1135 (2019)
- [30] Viazovetskyi, Y., Ivashkin, V., Kashin, E.: StyleGAN2 distillation for feed-forward image manipulation. *ArXiv abs/2003.03581*(2020)
- [31] Wang, G., Liu, Z., Hsieh, B., Zhuang, S., Gonzalez, J., Darrell, T., Stoica, I.: sensai: Convnets decomposition via class parallelism for fast inference on live data. *ML Sys Proceedings* (2021)
- [32] Wang, M., Deng, W.: Deep visual domain adaptation: A survey. *Neurocomputing*312, 135–153 (2018)
- [33] Wu, B., Iandola, F.N., Jin, P., Keutzer, K.: SqueezeDet: Unified, small, low powerfully convolutional neural networks for real-time object detection for autonomous driving. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 446–454 (2017)
- [34] Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR abs/1703.10593*(2017), <http://arxiv.org/abs/1703.10593>
- [35] Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., Stamminger, M., Nießner, M., Theobalt, C.: State of the art on monocular 3d face reconstruction, tracking, and applications. *Computer Graphics Forum* 37(2018)
- [36] Tero Karras et al, Alias-Free Generative Adversarial Networks (StyleGAN3), <https://arxiv.org/abs/2106.12423>, (2021)
- [37] B. Egger et al, 3D Morphable Face Models - Past, Present and Future, <https://arxiv.org/pdf/1909.01815.pdf>
- [38] <https://github.com/electronicarts/ReconstructionOfParametricFaces>: the supplementary codebase. (2022)
- [39] Lewis, John P., et al. "Practice and Theory of Blendshape Facial Models." *Eurographics (State of the Art Reports)* 1.8 (2014): 2.
- [40] Farahani, Abolfazl et al. "A Brief Review of Domain Adaptation." *ArXiv abs/2010.03978* (2021)

AUTHORS

Igor Borovikov received his Ph.D. in math at Moscow Institute for Physics and Technology in 1989. He is a Senior AI Scientist at Electronic Arts.



Karine Levonyan received her Ph.D. at Stanford University in 2019. She is an AI Scientist III at Electronic Arts.



Jon Rein graduated from Art Institute (CDIS) in 2004. Currently, he is a Senior Software Engineer at Electronic Arts



Pawel Wrotek graduated from Brown University in 2006 and received M.S. in Computer Science. He is currently a Senior Software Engineer at Electronic Arts



Nitish Victor received his M.S. from Rochester Institute of Technology in Game Design and Development in 2019. He is currently a Software Engineer II at Electronic Arts.



COGNITIVE GRAPHICAL PASSWORD BASED ON RECOGNITION WITH IMPROVED USER FUNCTIONALITY

Mozhdeh Sarkhoshi and Qianmu Li

Nanjing University of Science and Technology, Nanjing, Jiangsu, China

ABSTRACT

The fact that photos and graphics are more easily recalled by humans than text led to the proposal that visual passwords may be a viable alternative to text passwords in certain situations. User-friendliness characteristics of existing models are based on graphical password recognition, and the introduction of a new model that is related to the specifications and features of ISO standard usability and to the specifications and features of general usability specifications and features is being considered. Once these criteria and characteristics and sub-components of usability had been compared, additional usability features that could be included into the new graphical password model provided were discovered. There was a presentation of the graphical password model, which was separated into two sections, which included new users and current users. A questionnaire was used to evaluate the usability features and applicability of the prototype system after it had been implemented as a prototype system. After this step, the system was implemented as a prototype system and its evaluation and evaluation through a questionnaire was used to evaluate the usability features and applicability of the prototype system. Then there will be user input on the whole system, as well as the outcomes. The characteristics and specifications of the usability of the visual password prototype will be gathered and examined in this study. All of the percentages collected in this publication in connection to the findings and results from the point of view of usability are such that it is possible to conclude that the new visual password system is acceptable in its current form.

KEYWORDS

Graphic password, authentication, usability.

1. INTRODUCTION

In today's world, authentication methods, of which passwords are the most significant component, are extensively employed. As soon as the system determines that an input has the right username and password, the user is prompted to go through the authentication procedure. Users must initially register in order to be granted access to the system, and they must remember the username and password they created during the registration process in order to log in each time they wish to use it. Being aware of a user's password and username are the only things that can confirm their identity. Text passwords are often the sole method by which users are authenticated while accessing a network system. This approach is used by many networks, computer systems, and Internet settings today to verify their users' identities. Unfortunately, many passwords may be readily guessed or broken, making them vulnerable to attack. There are several and well-documented downsides of using this strategy [1].

Instead of using text-based passwords, visual encryption methods have been created as a replacement for them. As a result, picture codes have the potential to be more secure and reliable

David C. Wyld et al. (Eds): SIPP, NLPCL, BIGML, SOEN, AISC, NCWMC, CCSIT - 2022

pp. 17-24, 2022. CS & IT - CSCP 2022

DOI: 10.5121/csit.2022.121302

than earlier text codes since they govern the capacity to quickly detect and recall the image. Text-numeric coding has long been known to have a flaw, and this system is meant to solve that flaw. The premise that people recall images more readily than letters and numbers, and the concept that a picture is worth a thousand letters are both supported by certain research conducted by psychologists and software businesses. The fact that photographs and graphics are more easily recalled by humans than text led to the suggestion that visual passwords may be a suitable alternative to text-based passwords.

It is an authentication and validation system that operates on a graphical password interface, which allows users to pick certain photographs in a specified sequence using a visual password. Computer systems are routinely run under the control of an authentication system that relies on characters such as usernames and passwords. These sorts of systems have been known to include significant vulnerabilities; for example, users often use a basic password that may be quickly guessed by others. To be honest, it's tough to remember a password that's both safe and difficult to guess. Today, this approach is used as the authentication strategy for the vast majority of computer systems, whether they are network-based or host-based. A lot of people are already aware of the risks involved with this strategy. An assault linked to (Dictionary Attack), which is used by hackers to get access to numeric alphabetic passwords, is one of the most popular attacks in this field of study. However, there is no doubt that this form of assault is a very successful strategy. Because it just takes a short amount of time to figure out the password. Other disadvantages of utilizing this strategy include the same difficulty in learning and remembering passwords, since studies have shown that just a few kinds of passwords are simple for users to remember, as a result of which people often create and use the same passwords for all of their accounts [3]

As previously stated, the graphical password authentication system is a viable alternative to the numerical alphabetical form of password authentication. Specifically, this approach has been presented to remove the usual shortcomings and vulnerabilities of the primary methodology (numerical alphabetic technique). It is also possible that this technique will be better appropriate for producing passwords that are both more secure and more memorable for users. Users may recall photos more quickly than numeric alphabetic letters, according to one of the primary assumptions in this field. The other hypothesis is that images are worth thousands of dollars, according to the other hypothesis. These theories have been put to the test by the code, software firms, and some study in the area of psychology. Computer systems must meet certain security standards in order to function properly, which is particularly crucial given the proliferation of threats in this field. It's true that scientists and security professionals have long been concerned with the safety of computer systems, digital assets, and humans. However, security has been seen as a technological matter of considerable importance to date, and concerns have emerged in this respect. Users have resorted to passive or active usage of security technology in a variety of situations. Understanding may be vital in passive applications; nevertheless, in addition to active applications, consumers want additional usability elements as well as security-related solutions in passive apps. For example, they need characteristics such as expertise and mastery, simplicity of operation, ease of recall, contentment, and efficiency, among others.

Determining the authentication process is a procedure that decides whether or not a user is permitted to access a given system or piece of information. Despite the fact that passwords are still frequently used today to authenticate individuals, there are other options available, such as biometric systems and smart cards, that may be used instead.

Using these options, on the other hand, has a number of drawbacks. Biometrics involves a wide range of security concerns, many of which are connected to privacy; on the other hand, smart cards need a unique PIN in order to prevent them from being misplaced.

Consequently, passwords are still the primary method of authentication. A variety of shortcomings of traditional passwords are related to their usability, and these shortcomings are directly related to security difficulties. Users that are unable to choose secure passwords make the authentication procedure dangerous and provide possibilities for attackers to get access to credentials [4].

It is believed that passwords will be able to deal with two required requirements that are in conflict and conflict, and one feature that is extremely significant is that users will be able to passwords. There are various issues with passwords. Authentication processes must be conducted swiftly by users, but passwords must be safe, for example they must be secure, random and difficult to guess. Passwords must also be changed on a regular basis and in a secure manner. Users' preferences should evolve over time and not be the same for all of their accounts. They should also not be entered and kept in text files.

Utilitarian design is critical in the creation and development of an aesthetically pleasing graphical password system that also fits the demands and expectations of its users. ISO 241-11 defines usability as the degree to which a product may be utilized by individual users to accomplish their specified objectives in an effective and efficient manner, as well as adequately, in the appropriate area of application. Several academics have carried out investigations to propose new algorithms or enhance existing algorithms with the goal of boosting security and usability since the first graphical user authentication system was introduced by Blonder. Unfortunately, the majority of graphic password researchers have not paid attention to usability aspects. In most cases, researchers investigate graphical password security solutions, focusing on the probability of passwords failing during the authentication process, as well as user satisfaction and system operational aspects, among other things.

Do not put anything in it (especially the simplicity of remembering passwords). A critical topic to consider is the implementation of an entirely new graphical password system that offers a variety of interesting usability features.

2. DIAGNOSIS-BASED TECHNIQUES

Most articles from 2000 to 2015 have described that the methods available for diagnostic techniques are five designs. In the following section, existing methods will be reviewed and their strengths and weaknesses will be studied.

2.1. PASS FACE ALGORITHM

According on the idea that the human face is easier to recall than other photographs, Real User created a technology called Pass Face in 2002, which is now widely used. It is possible to pick images of previously seen people with this solution, which offers choices that prompt the user to select photos of previously seen people. When consumers have selected all four of their face photographs, the procedure is complete. Results of earlier research have indicated that users can remember their passwords more readily when using this approach as opposed to the text password method [6], despite the fact that it takes a longer time to login to the system when using this method.

In this approach, the user's gender, race, and face beauty are the three criteria that influence the choice of trend, allowing the selections to be predicted in advance of time. Although optional assignment makes the password more forgettable, it is given as a modification to make it more memorable. Other shortcomings of this technique are related to the processing time required. The

registration component of the Pass Face algorithm is the most time-consuming phase for users, resulting in a lengthier overall validation and authentication time than with a text password, according to the algorithm.

2.2. Already Seen Algorithm

This approach, which was first used in 2000 and reported by Lashkari and Farmand (2009), is given as a vast collection of photographs, some of which are random, and users are asked to choose a certain number of them from the collection. The photographs that were previously picked must be identified later on in order to verify the user's identity. When compared to text passwords and PINs, the login time is much longer, and 90 percent of users have been successful in this approach throughout the validation process, while other ways have only been successful in 70 percent of cases [7].

Some of this technique's shortcomings are discussed as follows. With traffic congestion and several photographs being given by the server, processing time will be prolonged; ii) Despite the fact that this solution has a lower password space, it will be more secure. The password formed is difficult to remember, iii) while the server must analyze documents pertaining to many users, picture selection is a time-consuming procedure, and iv) the overall time necessary to construct the password is prohibitively expensive. The time required for a text password is 25 seconds, but the time required for this strategy is 60 seconds.

2.3. Triangle Algorithm

This approach is based on the Shoulder-Surfing problem's graphical password-based solution. The system displays a number of items associated with passwords, and the user picks an area made by him. For instance, the system may show three items and the user may choose three things associated with the password, resulting in the formation of a triangle. By clicking on the inside of the invisible picture, the user may confirm. This technique is performed with the icons in various places on the screen [8]. Researchers recommend doing this technique numerous times to eliminate the chance of unintentional connection by clicking or rotation. As a result, the method's sluggish connecting procedure might be considered a significant shortcoming.

2.4. Picture Password Algorithm

This method is related to a graphic password scheme proposed by Sobrado and Birget, which is based on a visual password. Basically, this algorithm is designed for mobile devices such as PDAs. Password selection in this method is done using small photos in the form of themes provided in the form of cats and dogs or the sea and the beach. Therefore, users can be authenticated by identifying the viewed photos and touching them in the appropriate sequence using a stylus pen. Once the user can authenticate, he or she may change what theme or password. Researchers have also suggested that this process be repeated several times in order to minimize the possibility of connection in a random click or rotation mode [9].

Weaknesses: While this algorithm is provided with specific and specified photos, there will actually be a limited space to choose the password. In other words, as the researchers examining the algorithm have noted in their study, a numeric password is generated when each image represents a number of consecutive options. On the other hand, the selected sequences are shorter than the text passwords. One of the solutions offered to expand the password space created by using this solution is more complex and forgettable for users.

2.5. Story Algorithm

According to the Story algorithm, which was established in 2004, the user selects a set of photographs and is then tasked with identifying the required inventory from a collection of photos and images. These photographs depict locations, items, or people. To assess the users' ability, a series of nine photographs is supplied and the user is requested to choose four of them, while also providing a sequential component to aid with memory. The technology encourages that users construct a narrative to link their photographs and images [10]. According to the Monotheistic and Denominational Study (2009), users often forget their Story passwords and commit common blunders. Thus, in comparison to the validation of the Pass Face algorithm, the greatest shortcoming of this approach is the difficulty of recalling photographs.

3. PROPOSED METHOD

In this article, we propose a new model with high security for mobile. In this design, it consists of 6 dice, each dice containing 6 numbers from 1 to 6. Figure 1 shows the proposed scheme in this article. For example, suppose the suggested password is 1, 2, and 3, respectively. The user must move from dice to number 1 and swipe to dice number 2 and finally swipe to number 3.



Figure 1. proposed Cognitive graphical password

To set the password by the user, according to Figure 2, there are six dice with 6 numbers in front of the user. The user can select any combination of numbers from 1 to 6 and swipe in order. For example, suppose that the 1356 password is selected by the user and swipes from 1 to 3, then to 5, and finally to 6, respectively. This password is taken as a template in the publication. Figure 3 shows an example of entering the wrong password and two examples of the correct password.

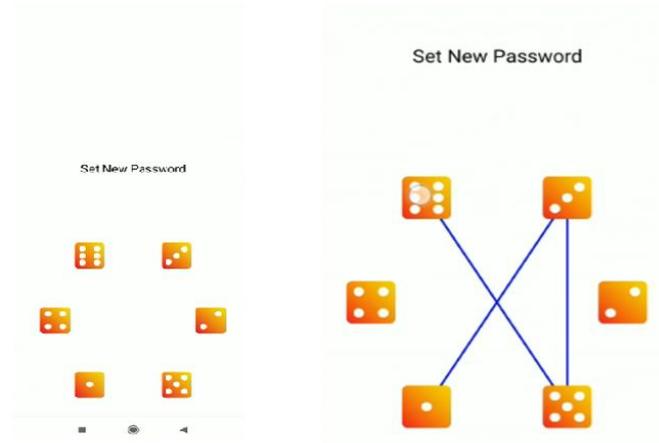


Figure 2. How to set a password in the proposed pattern

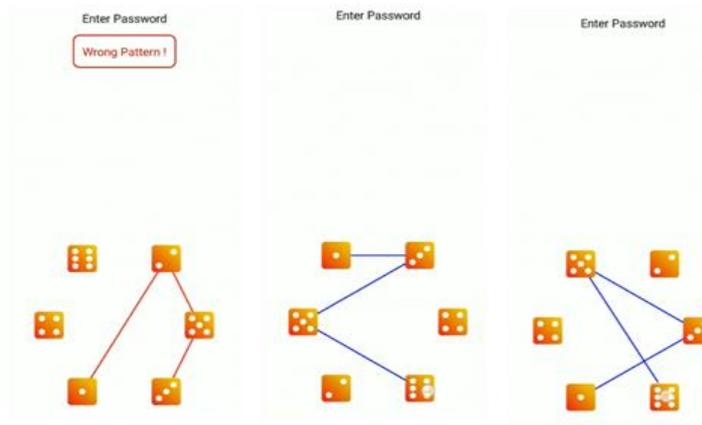


Figure 3. wrong and correct pattern

4. TESTING AND EVALUATION

The suggested system, a graphical password authentication mechanism, is assessed and tested by individuals who have used and tested it. Additionally, they will respond to questions in an online questionnaire designed to evaluate the system's usability qualities. The assessment plan that was created includes some information on the users. The participants are a group of 40 individuals, 15 of whom are female and 25 of whom are male. The participants were between the ages of 24 and 40. The participants were chosen as system users who accessed the system throughout the registration procedure and subsequently chose to connect. Finally, they answered to the online questionnaire's questions in order to provide a more detailed evaluation of the proposed system's usability qualities.

Participants tried to utilize the proposed authentication system by requesting the formation of a new user account and signing in after requesting the establishment of a new user account, as shown in the table below. As shown in the table below, the findings indicate that although all participants established their application accounts readily and without difficulty, they had divergent beliefs about how to connect to the system, as shown by their responses to the online questionnaire. The findings indicated that on the second day, 25 users successfully logged in on

their first try, 12 users successfully logged in on their second attempt, and just three users successfully logged in on their third attempt. On the third day, 20 users successfully logged in on their first try, 15 users received a success message on their second attempt, and 5 users successfully logged in on their third attempt. Table 1 contains all of the findings.

Table 1. Users trying to log in

	Create the user name and password			attempts to Enter the system		
	Attempt1	Attempt2	Attempt3	Attempt1	Attempt2	Attempt3
Day one- Creation	33	5	2	No need to enter the system		
Day2	No need for creation			25	12	3
Day3	No need for creation			20	15	5

Categories related to the structure of the questionnaire

Questionnaire review is the selected method, or in other words, the selected tool to verify the claims (objectives) of the research in terms of improving the usability characteristics of the graphic coding authentication system. 21 questions in 8 categories, which included some questions about the general information of users and the prototype of the system, as well as the usability features of the proposed system, formed the questionnaire form. The designed questionnaire, which included 21 questions in 8 different categories, can be seen in Appendix A.

- 1- General information
2. User comments
3. Evaluate the whole system
4. Evaluate the ease of use of the feature
5. Evaluate the simplicity of feature creation
6. Evaluate the simplicity of remembering a feature
7. Assess the simplicity of feature learning
8. Evaluate the outline

Finally, after analyzing all the answers, it is possible to realize that most of the system users have favorable feedback on all the usability features of the proposed method and also in evaluating the whole prototype of the system. All results are presented in Table 2.

Table 2. Results of the questionnaire analysis

Categories	Percentage
Evaluation of whole system	79.75 %
Easy to use	86.66 %
Easy to create	74.5 %
Easy to memorize	85 %
Easy to learn	81 %
Screen design	74.5 %

CONCLUSION

Based on the results in Table 2, this questionnaire explained that most users in general expressed their satisfaction with the usability features proposed in the proposed method and also with the whole prototype of the system as well as the visual design and outline. Have . The evaluation of the whole graphic password method had a percentage of 79.75%, which is very desirable, and it was determined that the system is generally acceptable to the participants in the dissertation testing and evaluation.

There was a variety of percentages between 74.5% to 86.66% in the usability characteristics of the proposed system. The highest percentage, which is equal to 86.66, is related to the simplicity of using the feature, which shows that the participants are satisfied with the simplicity of evaluating the system. The simplicity of creating a graphic password in the proposed design was 74.5%, so this favorable percentage showed that users created their password easily without any difficulty. Remembering the chosen password Despite 85% positive feedback showed that users can easily remember their passwords, due to the special features used in the design of this proposed design. 81% is related to the percentage of ease of learning the system, so this part of the questionnaire shows how easy it is to learn to use this system. The visual design of the proposed design and its overall design showed good results despite a positive feedback of 74.5%.

REFERENCES

- [1] S. Komanduri and D. R. Hutchings, "Order and entropy in picture passwords," in *Proceedings of graphics interface 2008*, 2008, pp. 115-122.
- [2] H. Gao, Z. Ren, X. Chang, X. Liu, and U. Aickelin, "A new graphical password scheme resistant to shoulder-surfing," in *Cyberworlds (CW), 2010 International Conference on*, 2010, pp. 194-199.
- [3] H. L. Arash, A. Abdul Manaf, and M. Masrom, "Security evaluation for graphical password," 2011.
- [4] N. Wright, A. S. Patrick, and R. Biddle, "Do you see your password?: applying recognition to textual passwords," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, 2012, p. 8.
- [5] Z. Erlich and M. Zviran, "Authentication methods for computer systems security," *Encyclopedia of information science and technology 2nd ed*, vol. 1, pp. 288-293, 2009.
- [6] S. Brostoff and M. A. Sasse, "Are Passfaces more usable than passwords? A field trial investigation," in *People and Computers XIV—Usability or Else!*, ed: Springer, 2000, pp. 405-424.
- [7] R. Dhamija and A. Perrig, "D'ej`a Vu: a user study using images for authentication," presented at the *Proceedings of the 9th conference on USENIX Security Symposium - Volume 9*, Denver, Colorado, 2000.
- [8] L. Sobrado and J.-C. Birget, "Graphical passwords," *The Rutgers Scholar, an electronic Bulletin for undergraduate research*, vol. 4, p. 2002, 2002.
- [9] W. Jansen, "Authenticating mobile device users through image selection," *The Internet Society: Advances in Learning, Commerce and Security*, vol. 1, pp. 183-194, 2004.
- [10] D. Davis, F. Monrose, and M. K. Reiter, "On User Choice in Graphical Password Schemes," in *USENIX Security Symposium*, 2004, pp. 11-11.

EXPERT SYSTEMS GENERATING MACHINE FOR IMAGE PROCESSING APPLICATIONS

Maan Ammar¹, Khuzama Ammar²,
Kinan Mansour³ and Waad Ammar⁴

¹Al Andalus University for medical sciences,
biomedical engineering, Al Kadmous, Syria

²Damascus University Hospital, Damascus, Syria

³Al Andalus University Hospital, Al Kadmous, Syria

⁴Zain Al Abedeen Hospital, Karbalaa, Iraq

ABSTRACT

We introduce in this paper what can be considered a new trend in expert systems field. It is generating different expert systems using the same software platform developed for this purpose, and called "Expert Systems Generating Machine for Image Processing Applications ESGMIPA". The machine is used to generate different expert systems in completely different application fields which indicates the feasibility of the proposal. Using what we called Domain Expert Guided Heuristic Search (DEGHS) and the machine, we generated an expert system that succeeded in cases where no algorithmic approach can be applied. Generating different expert systems using the same machine depends on the well-known fact that the function of an expert system is determined mainly by its knowledge base. The machine developed expedite very much the development of the expert system to reach best performance. The role of domain expert and the positive effect of the interaction between different domain experts in different fields is highlighted.

KEYWORDS

Expert systems generating machine, expert guided heuristic search, handwriting extraction, bacteria type automatic detection, bacteria colony image.

1. INTRODUCTION

In contrast to our previous research work presented in this conference last year where we used algorithmic approach to detect cancer in lungs CT images [1,2], this paper presents the ESGMIPA which is a software machine developed for generating expert systems in different application fields where the Domain Expert (DE) plays a central role in the generation process via what we called DEGHS. Two different expert systems in completely different fields are generated using this machine and the DEGHS search technique we developed. The first expert is used to extract unconstrained handwriting from unconstrained form bank checks, and the second one is used to automatically detect specific types of bacteria in microscopic bacteria colony image. Since the *DE plays the central role in the ES generation process*, the professionalism of this DE will appear in the design of the ESGMIPA, and in the generated Ess using this machine because the ES is *the executable version of the DE knowledge and experience*. Therefore, we will shed some light on the DE concept, and the professionalism level of the two DEs involved in reaching the ESGMIPA, and the way they developed their knowledge and experience.

1.1. Initiative, Experience, Knowledge and Domain Expert

The need and/or curiosity to know about some subject may lead to some initiatives in some field. The accumulation of results of initiatives will produce some knowledge in that field. If the acquired knowledge is worthy and extensive, then who made those initiatives becomes a domain expert.

A domain expert rarely starts from zero but builds on the knowledge of others who worked previously in the specific field. In most cases, study and training, as well as quality dedicated efforts by the person himself are very necessary to become a domain expert, like Medicine Doctor, Document analyst, geologist, and so on.

1.1.1. The Two Domain Experts involved in Developing the ESGMIPA

The two domain experts involved in developing the ESGMIPA are a DE in Electronic Circuits Design and Implementation (DEECDI) and a DE in signature verification and analysis (DESVVA). Both are the same person but in different age periods, and the final achievement reported in this paper came a result of the interaction between these two DEs.

In 1981, the first author of this paper became a DEECDI where he worked on developing ciphering/deciphering systems using pulse techniques in 1977 in the graduation project for the bachelor degree in Electronic Engineering [3], and completed a training course in electronic techniques at Dresden University, Germany in 1976 [4], and attended a very high level training course in Barry Research Corporation in the silicon valley of USA (California) on the design, operation and maintenance of a computerized sounder station for High frequency Communication for the Scientific Studies and Research Center (SSRC) of Damascus, Syria [5] after completion of a total immersion English language course at Berlitz school of languages, Palo Alto Office [6]. During his work at the SSRC he worked with professional Metal detectors [7] and Modems [8].

With the above qualifications, and during his work at the Faculty of Mechanical and Electrical Engineering at Damascus University as a lecturer assistant, he obtained a scholarship from the Japanese Government as a research student. Soon he was dispatched by Damascus University in 1983 to prepare for Doctor Degree in Information Engineering at Nagoya University, Nagoya, Japan. After finishing the Japanese language course, in a few days of work at Nagoya University Computation Center (a giant CC) with the signature data and the programs he received from his advisor, and with his professionalism and explorative mind in dealing with research problems as a DEECDI, he discovered the High Pressure Regions in off-line signatures by applying the half-power points of the curve of the resonant circuit from electrical circuit theory to the histogram of the signature and considered pixels with gray levels higher than this level as high pressure pixels and the others as low pressure ones. This principle gave amazing results in distinguishing between genuine signatures and skillfully forged ones, and the findings and their developments appeared in local (Japanese) and international publications [9-11]. During his work on computerized solutions of signature verification and analysis, M. Ammar studied famous references on suspect documents and their scientific examination [12] and linked between theory and application so that he became a domain expert in signature verification and analysis and in handwriting analysis. Later he was certified by the American board of Forensic Examiners (ABFE) as a *handwriting analyst* and a *forensic document examiner* [13,14], so that M. Ammar became formally a domain expert in signature verification and analysis (DESVVA). With the previous brief explanation, we can consider M. Ammar as a DEECDI, and a DESVVA. We notice that the first DE provided the second one with a golden chance to start his higher education in Japan quickly with a big momentum.

1.2. Related Works

Ammar M., et al, announced in 1985 reaching an automatic method to extract signature image from non-homogeneous noisy background as a part of a general approach to detect skilled off-line forgery signatures, which was unsolved problem [10]. Several months later they announced in TOKYO the complete method of automatic signature verification using pressure features in the monthly convention of the professionals in image processing and pattern recognition (Kenkyukai, held in Tokyo university in Feb. 1986) [9]. Later in October 1986, the topic was published in the 8th international conference on pattern recognition held in Paris [11], which means that the best specialists in the world have approved Ammar's method in High Pressure Regions extraction and using it for skilled forgery detection in off-line signatures, and M. Ammar became famous worldwide in this field. Due to the impressive content, another paper appeared on the same subject in a the (IEEE, Trans SMC journal) [15]. In Marse 1987, Ammar and his group presented the algorithm he developed to select the most effective features in a feature set of n features in $(n \times n)$ evaluations instead of $n!$ and used the results to divide the features into groups according to their type and effectivity [16]. In July of the same year in Montreal-Canada, another paper on the same topic appeared in a professional international symposium on handwriting and computer applications [17], followed by a detailed paper on the same topic in the book "computer recognition and human production of handwriting", world scientific [18].

In 1988, Ammar proposed the principle of signature description by computer which gives a symbolic description of the signature in a sophisticate manner and used it for the classification of a signature database into specific types and studying their nature. This work appeared in the 9th Int. Conference of Pattern Recognition in Rome, Italy [19], and used this description even for verification [20]. The application of signature description to signature analysis, announced by the same group, appeared in the 4th Int. conference of the Graphonomics society in Norway [21]. In 1989, M. Ammar, and as **a new trend in signature analysis by computer and its applications**, he developed an Interactive System with graphical and image display abilities, with the system ability to explain its response in natural language, for signature verification and analysis. He wrote a paper about the possible applications of this system with practical examples. One of the applications was to study the *stability of one's signature* (a common problem), and training those who complain the instability in the form of their signature to stabilize it, with some more applications, but warned that the same training application may be misused to produce undetectable forgery even by the best computerized systems. This paper appeared in the international conference of Image Processing and Analysis 7ICIAP in 1989 in Italy [22]. This interactive system received a great attention in Japan where it was written about and posted in a full page of the 17 million reader Japanese newspaper "Chunichi", and appeared in a televised report in the 6:00 PM prime time news of the TOKA television for 5 minutes. Later, I (Maan) recognized why the Japanese paid such attention to this system. The reason was because it appeared within the period of the National Project (the 5th Generation Computer) issued by Japan, which concentrate on developing the "intelligent computer". In fact, the interactive system, M. Ammar made, is really "**the truly intelligent system**" as described by Luger [23], and the computer running it is an intelligent computer in this field. In 1990 the detailed research about structural description and classification of signatures, appeared in a high rank and famous journal in this field [24]. In 1991 with the distinguished reputation Ammar realized, he was asked to analyze the documents of several actual cases of suspect Japanese documents. One among these cases involve 13 documents claiming over quarter million dollars. All these documents were judged by Ammar system to be forgeries. These findings appeared in a paper in an international workshop in Bonace, France, 1991 [25]. In 1992 M. Ammar, realized extraordinary results using a new technique he called "**Ammar matching technique**", and according to the results obtained using his data (prepared by Fujitsu company) he considered the performance a "**breakthrough in this field**". The new technique appeared in a paper in the proceedings of the 11th In. Conf. on

pattern recognition, held in the Netherlands [26]. Commenting on this copious production of signature related researches, R. Plamondon (*a prominent researcher in signature related field*) described Ammar M. and his group in his review paper [27] as **"the most active group in this field"**.

On a rather different track in the same field, in 1989, Ammar received an invitation letter from the International Academic Services (IAS) in the USA, congratulating him on his achievements and inviting him to work in the USA in research and teaching [28]. In 1990, and in connection with this letter, he travelled to the USA to communicate with those people, and to present his paper in the 10th Int. conference on pattern recognition 10thICPR held in Atlantic City. The paper was about a comparison between parametric and reference-pattern-based features in signature verification [29], which led to a well-known paper describing new advances in signature verification, by the same author [30]. After completing his mission in the USA, he decided to go back to his country and start *a new trip in the field of signature verification and analysis concentrates on building a PC-based signature verification and analysis software system on his cost*.

In 1990 he established a new research group in his country. They could build Personal Computer (PC)-based signature verification and analysis system (SIGVA 1.0) reported in an international conference in Canada in 1995 [31]. In 1995 also, M. Ammar received the certification of the *American Board of Forensic Examiners* in forensic document examination [13], and in forensic handwriting analysis [14]. In 1997 he received the certification of the justice ministry in Syria as the first examiner (highly qualified) of forensic documents [32].

Sigma-1.0 attracted investors from Germany and USA to Syria. The negotiations led to cofounding with Sam Koo ASV Technologies Inc. in 1998 in USA. The work continued in developing the ASV system for USA banks by ASV Technologies Inc. via three groups: the first and essential one in Damascus, the second one in Stuttgart in Germany and the third one in New York in the USA with the supervision and coordination between the three groups by M. Ammar until the first ASV (Automatic Signature Verification) server was set up in NY in 2000, and the US patent of that system was received in 2002 [33].

In 2001, M. Ammar joined Applied Sciences University in Amman, Jordan as full professor specialized in Image processing and Intelligent Systems. While teaching Image processing, Artificial Intelligence, Decision support Systems, and several other subjects, he published several papers in the fields of AI, Computer Vision, and Image processing, *with some relation to the content of this paper* [34-37].

In 2010, he received an invitation from Lambert Academic Publishing (International Publisher) to write his experience and works in a book. In 2011, the book "Intelligent Signature Verification and analysis" was published by the LAP [38]. With the progress of the work of ASV Technologies, serving hundreds of banks in the USA, more and more requirements appeared. Among them increasing further the correct verification rate of the system. As a response for that need, Ammar proposed and implemented the "multi-feature set" verification decision which gave important improvement in correct verification rate [39-40]. By that time, the system had verified over one billion bank check without wrong decisions, with moving a handful of signatures at the end of each batch of tens of thousands of signatures as suspects for visual verification [39].

At this point in the trip of developing signature verification and analysis software, the company (ASV Technologies Inc.) asked M. Ammar to work on handwriting extraction from bank checks. This request led to the achievements reported in this paper.

2. THE COMPLEXITY OF EXTRACTION OF UNKNOWN HANDWRITING FROM UNKNOWN CHECK IMAGE

After discussion with the company about what is really wanted, the result came as follows: The wanted work is extracting “unconstrained handwriting on a bank check image” (may take any form), and the check design is also “unconstrained”, (the check may come from different banks), and consequently, the design of the check is *unpredictable*. Moreover, all check images are binary, and some are with bad quality.

Taking in consideration that the bank check is a complicated design in nature (contains different fields for writing and signing, symbols, decorations, logos, etc.), everything could be variable, and the objects we should extract “handwriting” are variables and unconstrained. For the first moment, the task seemed to be extremely difficult (if it were possible at all), but finally, M. Ammar accepted the challenge. The research achievements are presented in this paper with some further developments that led to the ES generating machine, reported in this paper. The four bank checks shown below in Fig. 1 are examples of the data we must deal with. With this kind of problems, Expert systems could be the suitable approach, therefore, we will explain in brief about an expert system and how to build it, then apply that to our problem.



Fig. 1 (a, b, c, d, raster) Four examples of the bank checks to work with.

2.1. What is an Expert System?

An expert system can be defined in different ways, and there are many kinds of block diagrams explaining its function, depending on the field and the case dealt with, however, in general, any expert system must contain at least: (1) a knowledge base, (2) an inference engine and (3) a user interface.

2.2. Definition an Expert System

Simply speaking, it is a computer program using artificial intelligence techniques. It uses a database of expert knowledge to offer advice or make decisions in some specific area (the area

here is handwriting and bank checks environment). Concerning a general block diagram, we will adapt here the block diagram suggested by JA Bullinaria, 2005, [41] shown in Fig. 2, because we found it informative and suitable. The heart of this expert system is the knowledge base (facts, rules and heuristics) and the inference engine that applies rules to the facts to infer new facts. Rules that represent the knowledge of the *domain expert* are collected and formulated by the *knowledge engineer* and programmed and stored in the KB by the specialized *programmer*. The domain expert, the knowledge engineer and the programmer are the necessary team to build the ES.

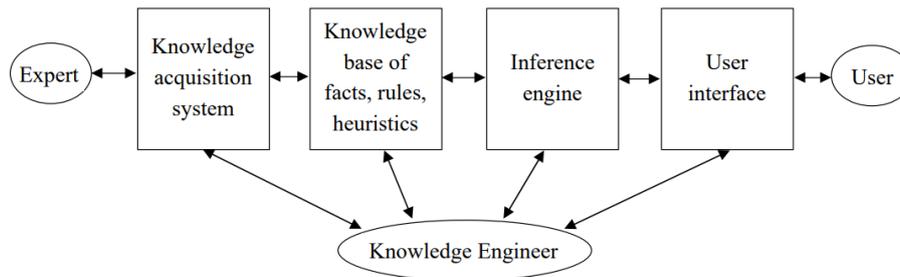


Fig 2. A general block diagram of the expert system.

2.3. Considerations in Forming the ES Building Team

The team of building an ES must be formed taking in consideration the following points:

- 1- The Domain Expert (DE) is a person with a professional experience and knowledge in the specific domain.
- 2- The Knowledge Engineer (KE) is a person who is familiar with the specific domain and with programming. He must understand the key concepts from the DE and formulate them in a form the programmer can understand and program in executable form.
- 3- The programmer (P) is a person who masters programming in the desired language.

Now, we reach an important question: *is it necessary to have three persons to form the team and build the ES?* The answer is: in the general case, Yes but in some cases, No, we do not need three persons. For example, in reference [22] the DE, the KE and the P were the same person who built the interactive ES. In a case like this we can get the highest efficiency. In our case in building an ES for handwriting extraction from bank check environment, the DE and the KE are the same person (M. Ammar), and the programmer is a person professional in programming in Matlab and C++. In the expert system that detect the specific bacteria in a microscopic Bacteria Colony Image (BCI), we had to have three different persons as a team to build the ES. Now, we will characterize our problem viewed by the DE.

3. CHARACTERIZATION OF THE PROBLEM

As we mentioned above, the check form is unconstrained so that logos, decorations, symbols, and other objects may differ from bank to another as we have shown in Fig. 1. All what we know is that the objects we must extract are “**handwriting**” which may contain names, numbers, symbols and words. We must keep in mind also that the handwriting style may differ substantially for one person to another. We will try now to characterize the wanted objects in general:

- 1 – It is a pen line produced by freehand movement controlled by the brain.

- 2 –It almost may never contain completely straight lines or 90-degree angles.
- 3 – It is Smooth and should not contain broken strokes, unless it is a forgery.
- 4- Signature which must be on the check is a *special type of handwriting* may contain decorations and special long curved strokes.

We will build a Rule – based ES *to extract handwriting from bank checks*. The contents of the KB are facts, rules, and heuristics, summarized as follows:

Facts:

Facts related to this problem domain are features of the signature in special and the handwriting in general. Those features are studied, extracted and used extensively [10-11,16-22,24-26, 29-31].

Rules:

The rules will depend essentially on the ranges the values those features may vary inside and still differ from printed objects. Unlike signature verification cases in which those ranges can be learned from the training data, here, those ranges will be found heuristically with the help of the DE and the software machine designed for this purpose. In our problem, there is no training data, but there is only test data.

Heuristics

Instead of “state evaluation function” used in heuristic search in the search process to reach the goal state in a problem like chess game [23], we will use here what we called *Domain Expert Guided Heuristic Search* (DEGHS) because the evaluation of the distance to the goal (best result here) can’t be estimated by a number, but it is a visual judgment of a general view of the handwriting on a check. The DE evaluates the improvement obtained and then select the new movement (changing range values, introducing new features, etc.).

4. HANDWRITING EXTRACTION APPROACH

This approach consists of 2 main stages: (1) preprocessing, and (2) handwriting extraction.

4.1. Preprocessing

Before starting the actual handwriting extraction process, we segment the check image into components in order to extract features from these components (Facts), and then apply DEGHS using the rules suggested by the DE. Preprocessing is done in three steps:

1. Applying a Connected Component Labeling (CCL) process to the check image.
2. Closing using a square structuring element with 3 side value.
3. Dilation using a square structuring element with 4 side value.

This preprocessing fattens the printed characters giving them higher density to be removed later by rules.

4.2. Handwriting Extraction

As we mentioned in section 3, we *do not know anything about handwriting on the binary check image except that it is handwriting* (no information about spatial positions, form or density). We

also don't know anything about the design or content of the bank check therefore we will detect the handwriting by applying this global rule: delete any object on the check image if it has any one of the non-handwriting characteristics. The remaining will be the handwriting, *if available*. This solution is very general. We have to go inside its specifics. We will approach this solution as follows:

4.2.1. Approaching the Solution

We started from this fact: a human can recognize handwriting from printed text and other shapes in a document easily. *If we asked a DE, how can a human do that? His answer might be:*

Because the difference in the general appearance of handwriting described in section 3 four points and the general appearance of the printed text, other shapes and symbols is very clear.

Of course, this is a general answer. If we asked him to be more specific, his answer might be: *because the printed text features like the compactness of the characters and sharp change in stroke direction, and so on., are clearly different from those of handwriting already explained in section 3.*

Now, we are at the starting point of sketching a solution. Our DE (M. Ammar) who is a DE in both handwriting analysis and computerized solutions related to signature in special, and handwriting in general, will suggest the ***requirements of the solution.***

4.2.2. Requirements of the Solution

- 1- We need some features to be used for distinguishing between handwriting objects and other objects might find in the check. Essentially, those features should be available in the following references [10-11, 16-22, 24-26, 29-31], as mentioned before.
- 2- The performance of these features must be evaluated with some data to choose the suitable ones.
- 3- Since there is no training data for the contents of the design of the checks or for the handwriting, we have to proceed as follows: (1) select some features recommended by the DE based on his knowledge and experience, (2) start testing with some heuristics suggested also by the DE about the ranges of the values of the features that can be used to distinguish between handwriting and other objects, (3) update the ranges of values and/or used features according to the results obtained so that better result is hoped, (4) retest and evaluate the results. (5) repeating steps 2-4 until we get the best result.
- 4- How can we evaluate the result? When using heuristic search in AI problems like chess playing, there is a state evaluation function that return a value telling us how far from the goal, and based on that value the next move is estimated. Here, we can't design such cost function because the distance to the goal can't not be measured by numbers because evaluation is visual. Now we must define the goal state and in between states.
- 5- Our goal state is a check image contains only handwriting. ***Of course, this is the optimal case.*** This case might be impossible to reach because of the overlapping between handwriting objects and others, however it can be approached. Therefore, our goal state is the best state (bs) in which the maximum amount of handwriting extracted with other objects removed. This state can't be measured by numbers but by **visual estimation** given by the DE, therefore we call this kind of heuristic search "Domain Expert Guided Heuristic Search" (DEGHS).
- 6- We mentioned several times the terms "handwriting objects" and "other objects" therefore, the first step we must go is segmenting the check image into its objects. this can be done by Connected Components Labeling (CCL).

- 7- In order that the DE can give his response flexibly and in reasonable time, we have to provide him with these abilities:
- 8- Displaying the input image, (2) displaying the processed image at any stage, (3) displaying any component selected with its image, some other helpful images, and values of all features used, as well as any preprocessing done with parameter values used. The DE must also be provided with the ability of easy selection of any segmented object (CC) using its label, for convenience. We must also provide the ability to select the features the DE wish to use with their value ranges and any conditions desired (AND, XOR, etc.).
- 9- When we started to work with this subject, we designed and implemented a platform that enable the DE to interact easily with all what we mentioned above (with the help of the KE and the P, if needed) to give him high flexibility in suggesting heuristics, testing them, evaluating the result, making changes and retest again and again until the best result is reached. Fig. 3 shows this platform during one of the DE tests. This platform with the software tied to it is called ESGMIPA.

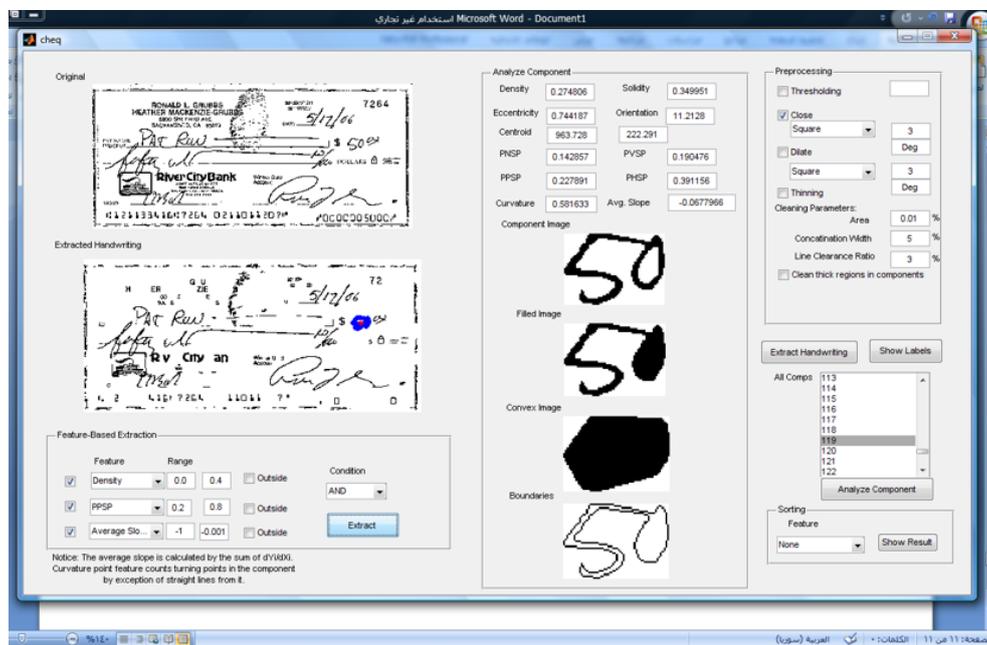


Fig 3. a screen shot of the platform (ESGMIPA) during a specific stage of the DE tests.

4.2.3. Description of the Screenshot

The image in the top left corner is the original image; the image below it directly is the result of an intermediate stage in which the component No. 119 which is actually the handwritten number "50", and appears in the right side of the check image colored blue for easy location by eye. In the vertical field in the middle in which the image of the No. 50 appears, three more images helpful in evaluation: filled image, convex image, and boundary image. Above the component image, the values of 12 candidate features of the component appears (density, eccentricity, centroid, PNSP, PPSP, curvature, solidity, orientation, PVSP, and PHSP). In the rightmost field on top, the possible preprocessing operations, some special cleaning operations, below that in the same field the number of the CCs appear with the ability to select any component and see all relating results like those appear in this screenshot. The ability of sorting the CCs is also provided. Going back to the left vertical field and below images, we find adjustable feature value ranges with some logical conditions to apply.

4.2.4. Obtained Results using the Platform (ESGMIPA) for ES Development

Fig. 4 (left) shows the handwriting extracted from the binary check image shown in Fig.1 (b), Fig. 4 (right) shows the input binary image without handwriting obtained by ANDing the complement of the extracted handwriting image with the original input image. Fig. 5 (right) shows another example of extracting handwriting from the check shown in Fig. 5 (left). The ES was tested with tens of checks (83 check images) and gave very good results.

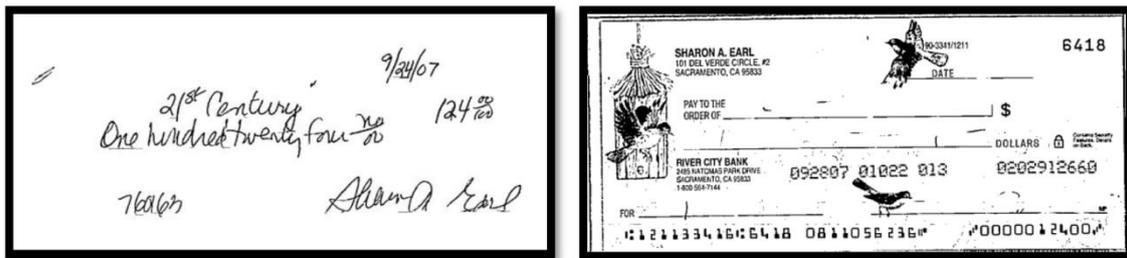


Fig 4. Extracted handwriting (left) and input image without handwriting (right).

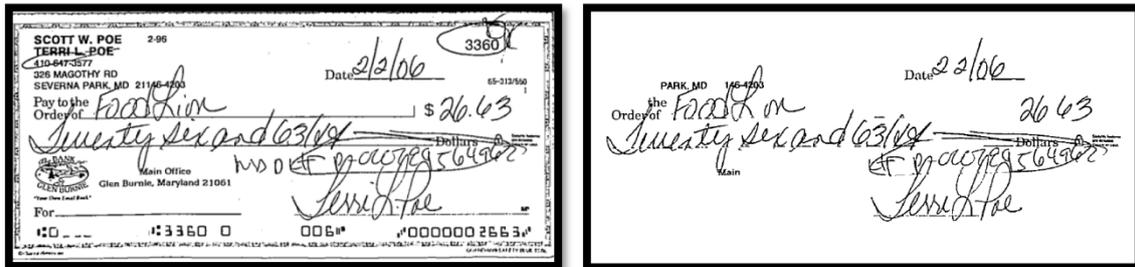


Fig 5. original binary image (left), and extracted handwriting (right).

4.2.5. The Extreme Cases

We may consider the checks shown already in Figs. 4 and 5 ordinary cases of bank checks written with full spontaneity, as ordinary cases. Now what happens if the check image can be considered to have unexpected content (no handwriting or full of noise, as the cases c and d in Fig. 1.). The result of these two checks is shown below in Fig. 6. As can be seen, we got almost complete performance where in the no handwriting check we detected no handwriting, and the heavy noise in the noisy check was removed without effect on the handwriting detected completely, with only on printed letter (can be removed by post processing).

At this point of development, the company asked whether the system can extract Chinese handwriting and sent us a test check shown in Fig. 8(a). We started to work with this check using our ESGMIPA, and displayed the number of the connected component (object) directly beside it as shown in Fig. 7. This way of display is very handy and gives us better understanding of the image components at a glance. We could in a few hours modify the content of the KB of the ES to get the result shown in Fig. 8 (b). The input image without handwriting by ANDing is shown in Fig.8 (c).



Fig 6. (a, b, c, d, e, f, raster order): a: a check without handwriting, b: the result of handwriting detection where nothing detected, c: the original check by ANDING, d: a noisy bank check, e: detected handwriting, f: original check by ANDING. (almost complete performance).

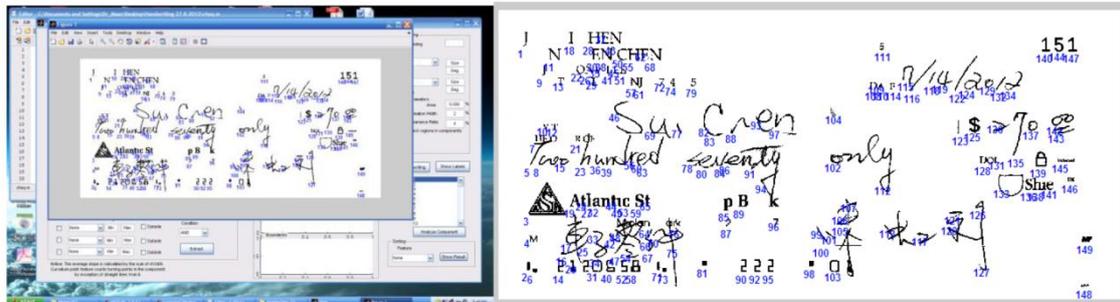


Fig 7. displaying the number of the CC directly beside it.

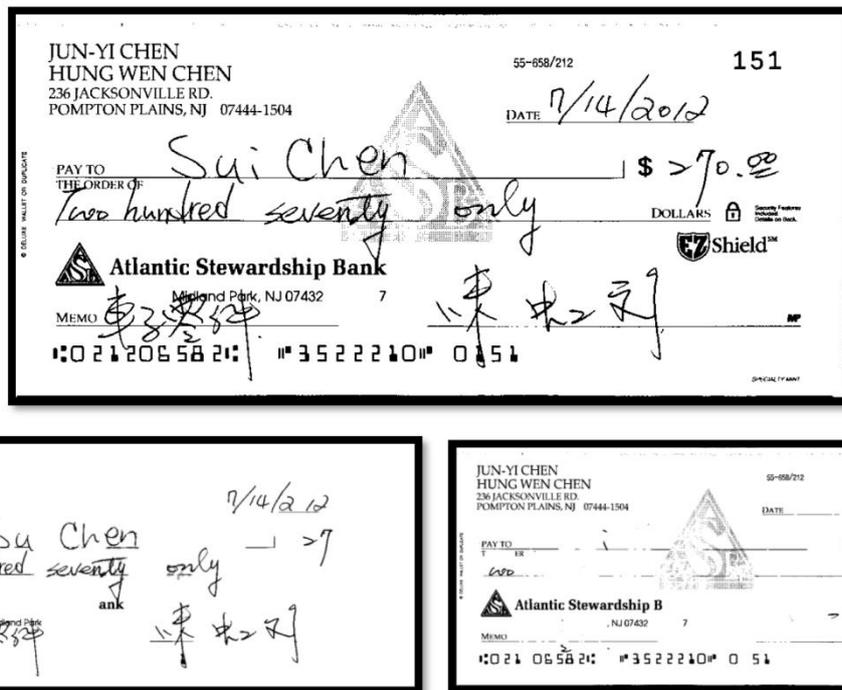


Fig. 8 A check image with Chinese handwriting (top), extracted handwriting (bottom left), the check without handwriting by ANDING (bottom right).

AS can be seen in Fig. 8, the result is excellent.

5. REINFORCING THE PRINCIPLE OF ESGMIPA

In fact, the principle of “expert system generating machine” was proposed by the first author (M. Ammar) in a local publication at Damascus University 9 years ago [42]. In this section, and to *reinforce this principle*, we introduce using our ESGMIPA to generate another ES to detect specific types of bacteria in a BCI.

5.1. An ES for Detection of Specific Bacteria Types in a BCI

Actually, M. Ammar is a DE in Biomedical Engineering (BE) field also, since he has been a teaching staff member in the BE department at the FMEE, Damascus university for 32 years, and served as Head of Department for 8 years. During this period, he translated and wrote several books for the department [42-45], and was active in interaction with Damascus hospitals and the department, especially with Damascus University educational hospital. In that hospital one of the coauthors is working as head of the *Bacteria Laboratory*. During discussion with her, it appeared that detecting automatically by computer some specific bacteria objects in a microscopic BCI image containing very big number of objects like that shown below in Fig. 9 is highly desired. Therefore, we found it a good chance to test the ability of our machine to generate expert systems in a field completely different from checks and handwriting.

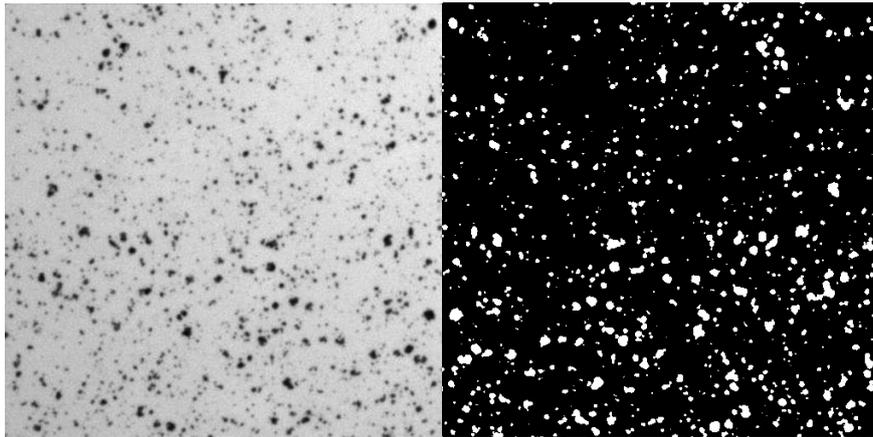


Fig. 9. Bacteria Colony gray image, and its thresholded version.

The DE and the KE made a discussion to characterize the problem before asking the programmer to program a system for that purpose, of course using the ESGMIPA. The problem here is rather easier than handwriting problem because there is no overlapping between objects.

5.2. Steps used in Developing the Bacteria ES

1 -Specifying the general knowledge in hand

1. The image to work with is a gray level one, shown in Fig. 9.
2. The image contains objects darker than the background (lower gray levels).
3. The image contains very big number of objects vary in shape and size.

2 – Getting more specific knowledge the KE derived from the DE

1. The objects (bacteria) to be detected are among the larger objects in the image, and they are approximately bigger that 5% of the area of the total image.
2. Objects to be detected are two types:
3. The first type has relatively low density measured by the area of the object divided by the area of the rectangle confining it.
4. The second type has a higher density compared with the first one.
5. The circumference of the objects of the second type is smoother than the first one and more homogenous.

As we can see, this knowledge is general and relative. The exact knowledge of aimed objects is known only by the DE, therefore we *must follow the heuristic methodology in cooperation between the DE, the KE, and the programmer, using our ESGM*.

3 – Specifying the heuristic methodology (HM) to be used

- 1 – Design and/or select suitable features to detect approximate shapes of the objects according to approximate knowledge described in the above 5 points.
- 2- Using this approximate knowledge to retrieve objects satisfying its content, (candidates), and show them to the DE.
- 3 – Modify the features and/or their values according to the comments of the DE to become closer to detect the wanted objects.

- 4 – Repeating 2 and 3 until reaching the goal which is (detecting the desired bacteria objects as accurate as possible, if exist).
- 5 – The final finding of the features, their values, rules and conditions become the content of the KB.

4, Applying the actions needed to implement the HM

- 1 – Segmenting the binary BCI into its components in which we must search for the bacteria objects to be detected.
- 2 – Deleting the objects with area less than 5% of the total image area.
- 3 – Designing or selecting a function to compute the density and another one to compute the smoothness and then fine tune their parameters to reach the goal with the supervision of the DE
- 4–Determining the logical functions necessary to combine the effects of the functional functions to reach the goal.
- 5 – Using DEGRHS to reach the minimum and maximum limits of the features values and the necessary logical operations to give the final form of the Rules to be used by the ES to efficiently detect the wanted objects (bacteria).

We will show below the results of some key actions implemented to reach the goal:

- 1 –Thresholding the BC image. The result is shown in Fig. 9.
- 2 -The result of deleting objects with area less than 2% of the total area of the BC image (here, although the estimated area of the objects is around 5%, the team preferred to see all objects above 2% first appearing in Fig. 10).

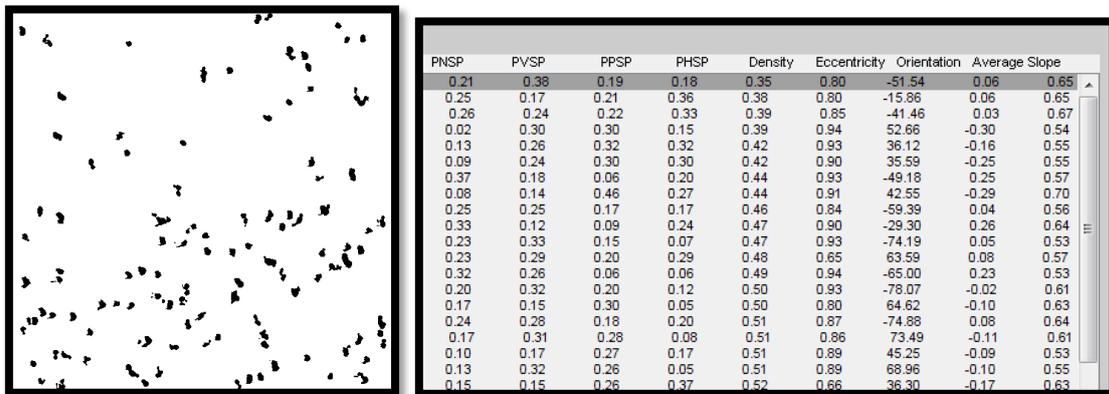


Fig. 10. Objects with area larger than 2% of the total BCI area (118 objects) and the values of 9 features computed for them shown in the table.

A part of a table of 9 features of the 118 components is shown in Fig. 10. The features are: Curvature, average slope, orientation, eccentricity, density, percentage of positively, negatively vertically and horizontally slanted pixels in the boundaries of the object (CC). Features in the table are sorted according to the values of “density” feature. Fig. 11 shows the result of deleting components with area less than 5% of the BCI area. Remaining components are 24 and the Values of the 9 features of all the 24 components (objects) sorted by the curvature (the right most column in the table) are shown in the table in Fig. 11.

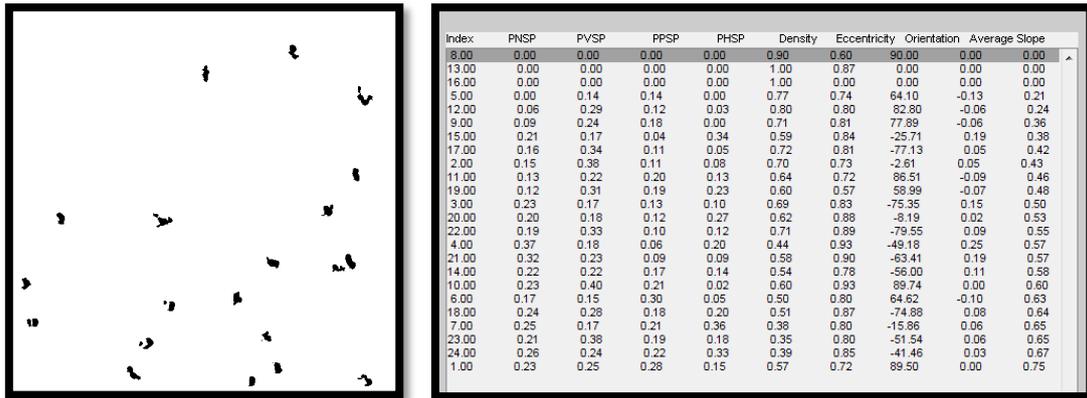


Fig. 11. The remaining 24 objects and the values of the 9 features.

Fig. 12 (left): shows the 12 objects remaining from those in Fig. 11, (middle): the Table of the nine features sorted by values of curvature in the rightmost column, (right): remaining 5 objects belonging to the values (0.56-0.67) in the right most column. These 5 objects accepted by the DE as type I bacteria objects are shown enlarged in fig. 13.

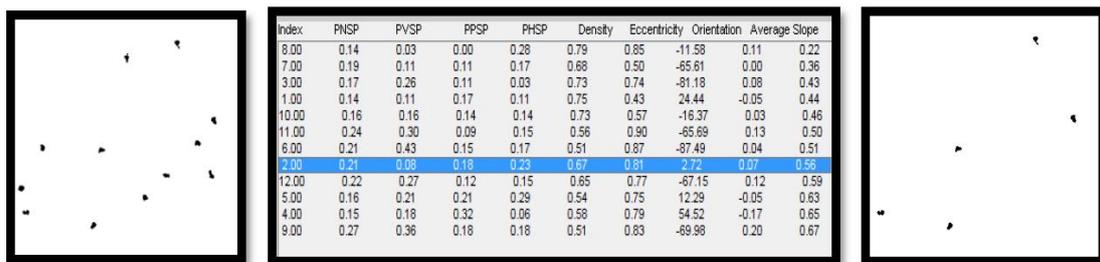


Fig. 12 (left) remaining 12 objects remaining from those in Fig.10, (middle) Table of the nine features sorted by values of curvature in the rightmost column,(right) remaining 5 objects belonging to the values (0.56-0.67) in the right most column. These 5 objects are shown enlarged in fig. 13.

5 - Bacteria type I final result



Fig. 13. The five objects of type I bacteria reached at the end of the DEGHS.

6 - Bacteria type II final result

Fig 14 shows the two objects of type II bacteria reached at the end of the DEGHS (bacteria objects with higher density, higher smoothness, and more homogeneity) with their geometrical measures. The enlarged objects appear in 4 types: original, filled, convex hull and borders represented by 8-directionals. These figures with the tables are used in evaluating the results during DEGHE. Fig. 15 shows the two objects as a final result.

Any more development?

When introducing a method or approach, especially if it is new, this question appears: is there any limitation in performance? Here is our answer: as far as realizing the principle of generating expert systems in the sense we proposed is concerned, we see no limitations, however, if the volume of data used in test is concerned, we believe that we have to test our machine with much more samples for English language checks, especially with the extremely variable environment (unconstrained handwriting and check background design). *Tens of samples are not sufficient to judge completely the performance.* In this regard, we may say, when the volume of checks is very big, we may use “grouping” principle to expand the ability of the system. In fact, “grouping” is one of the secrets of success of ASV Technologies over 20 years of work without problems with millions of checks investigated every day. Concerning the Chinese language, we tested only one sample with excellent result, but it must be tested with sufficient number of samples. Finally, we would like to say: *we concentrated in this paper on the success of the principle, in general, not on the fine details.*

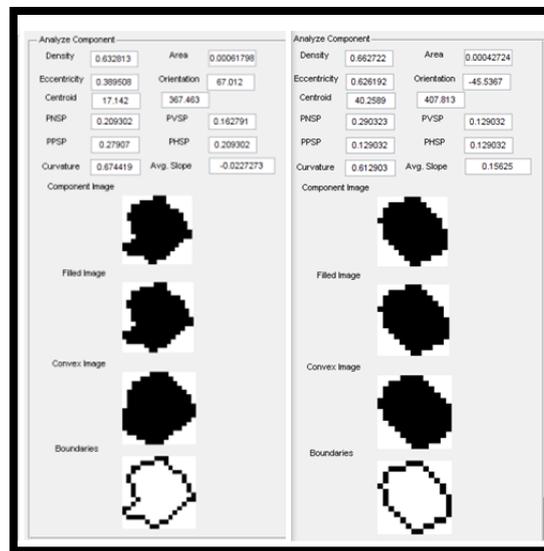


Fig. 14 The two objects of type II bacteria reached at the end of the DEGHS with their 4 shapes and values of candidate features.



Fig. 15 the final result of type II bacteria objects at the end of the DEGHS.

6. CONCLUSION

WE have introduced in this paper what we called " Expert Systems Generating Machine for Image Processing Applications (ESGMIPA). This machine is a software one designed to give the Domain Expert who will guide what we called "Domain Expert Guided Heuristic Search" the widest choices of processing the image, computing the values of its features, displaying some useful types of its images, enabling him to apply some logical conditions (AND, XOR, ..) to the features when applying the Rule of the ES to be generated, to solve the problem in hand, using some possible preprocessing techniques, and showing all these choices and their results in one screen giving him the ability to evaluate the situation at a glance, and giving his judgement to proceed to a next move or stop and accept the final result. Besides reaching a solution to some problems where no algorithmic approach can be applied, We found by practical applications that this machine speeds up very much reaching the desired solution for this class of problems. As a real application, we applied the machine to generating an expert system to extract unconstrained English handwriting from unconstrained form (design) binary bank check with high effectivity, even if the check is noisy sometimes. We also could modify the KB of the first ES quickly to do the same thing when the language is Chinese. As reinforcing of the principle of ES Generating Machine, we could easily generate an ES that detects the objects of two types of bacteria in a bacteria colony microscopic image containing very large number of microscopic bacteria objects, efficiently. We showed also that interaction between different DEs may be very useful in research.

ACKNOWLEDGEMENTS

We wish to thank ASV technologies Inc. and Damascus University educational hospital for providing us with the data and Al Andalus University for various types of support.

PARTICIPATION OF AUTHORS

Maan Ammar: main researcher, Khuzama Ammar: Domain Expert, discussion and testing, Kinan Mansour: Domain Expert, discussion and testing, Waad Ammar: Domain Expert, discussion and testing.

REFERENCES

- [1] M. Mounief, et al, Automatic Detection and Extraction of Lungs Cancer Nodules Using Connected Components Labeling and Distance Measure Based Classification, 9th International conference on signal, image processing and pattern recognition, pp.41-53. Vancouver, Canada, 2021.
- [2] M. Ammar, et al, Using Distance Measure Based Classification in Automatic Extraction of Lungs Cancer Nodules for Computer Aided Diagnosis, Signal & Image Processing International Journal, Vol 12, No. 3, pp. 25-43. 2021.
- [3] M. Ammar and H, Alodda, A Telephone calls ciphering/deciphering system using pulse techniques, TR E/3/1977, FME1977.
- [4] M. Ammar, A certificate of completion a training course in electronic techniques from July 20, 1976 to August 20 1976, issued by Dresden technical University.
- [5] M. Ammar, Certificate of completion a BR Communications course in VOS1 sounder operation and maintenance, issued at Barry Research Corp, Sunnysvale, USA, issued at 11th day of March 1978.
- [6] Ammar, Certificate of completion a "total immersion English language course", Dec. 12 1977 to Jan11, 1978, issued at Palo Alto office, in Jan. 11, 1978.
- [7] Ammar M., A professional metal detector (deep seeker) for ferro/non-ferro metal objects, TR E/04/79, SSRC, Damascus, 1979.
- [8] M. Ammar, 2400 bps modem for telephone line and wireless transceivers, TR E/22/5/1979, SSRC, Damascus Syria.

- [9] M. Ammar, Y. Yoshida and T. Fukumura, Automatic Off-line Verification of Signatures Based on Pressure Features", Institute of the Elect. and Communications Eng. of Japan (IECEJ), PRL-85-37, Vol.85, No.173, pp. 23-34, Oct. 1985.
- [10] Ammar M., Yoshida Y. and Fukumura T., Automatic Extraction of Signature Image from Handwritten Documents, National convention of the IECEJ, Yokohama, Japan, p.135, 1985.
- [11] M. Ammar, Y. Yoshida and T. Fukumura, A New Effective Approach for Automatic Off-line Verification of Signatures by Using Pressure Features, IEEE Computer Society, Proceedings of the 8th Int. Conf. on Pattern Recognition, Paris, P. 566-570, Oct. 1986.
- [12] W. R. Harrison, Suspect documents-their scientific examination, Universal Lexis Nexis, 1977.
- [13] M. Ammar, Board Certificate in Forensic Document Examination, American Board of Forensic Examiners(ABFE), Certificate NO. 07,1993.
- [14] M. Ammar, Board Certification in Forensic Handwriting Analysis, American Board of Forensic Examiners (ABFE), Certificate NO. 03,1993.
- [15] Ammar M., Yoshida Y. and Fukumura T., Automatic off-line verification of signatures based on pressure features, IEEE Trans on Man Systems and Cybernetics, Vol SNC-16, No 3, pp. 39-47, 1986.
- [16] Ammar M., Yoshida Y. and Fukumura T., Features extraction and selection for simulated signature verification, the summit of professionals in image processing and pattern recognition (Kenkyukai), IECEJ, Tokyo, Japan, IE86-120, pp. 81-88, 1986.
- [17] M. Ammar, Y. Yoshida and T. Fukumura, Feature Extraction and Selection for Simulated Signature Verification, Proceedings of the 3rd Int. Symposium on Handwriting and Computer Applications, Montreal, Canada, pp. 167-169, 1987.
- [18] M. Ammar, Y. Yoshida and T. Fukumura, Feature Extraction and Selection for Simulated Signature Verification, Computer recognition and human production of handwriting, R. Plamondon, et al., world scientific, pp. 61-76, 1989.
- [19] M. Ammar, Y. Yoshida and T. Fukumura, Description of Signature Images and Its Application to Their Classification, Proceedings of the 9th Int. Conf. on Pattern Recognition, Rome, Italy, pp. 23-26, September 1988.
- [20] M. Ammar, Y. Yoshida and T. Fukumura, Application of signature description to verification, Technical Report, IE/YL/3/1988, Faculty of Information Engineering, Nagoya University, Japan, 1989.
- [21] M. Ammar, Y. Yoshida and T. Fukumura, Signature Analysis by Computer, Proceedings of the 4th Int. Graphonomics Society Conference, P. 56, Trondheim, July 1989.
- [22] M. Ammar, Applications of Signature Analysis by Computer and the Consequence of its Possible Misuse, Proceedings of the 5th Int. Conference on Image Analysis and Processing (5ICIAP), Positano, Italy, world scientific, pp. 535-542, Sept. 1989.
- [23] G. Luger, Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 6th Edition, pearson, 2008.
- [24] M. Ammar, Y. Yoshida and T. Fukumura," Structural description and classification of signature images", Pattern Recognition Journal, Vol. 23, No.7, pp. 697-710, 1990.
- [25] M. Ammar, Identification of fraudulent Japanese signatures from actual handwritten documents: A case study, Proceedings of the Second Int. Workshop on Frontiers in Handwriting Recognition, Bonas, France, 1991.
- [26] M. Ammar, Elimination of skilled forgeries in off-line systems: a breakthrough, proceedings, 11th Int. Conference on Pattern Recognition, the Netherlands, IEEE computer Society, pp. 415-418, Sept., 1992.
- [27] R. Plamondon, Progress in automatic signature verification: the state of the art—1989–1993, Int. Journal of Pattern Recognition and Artificial Intelligence, Vol. 08, No. 03, 1989-1993 (1994).
- [28] M. Ammar, Invitation letter received from International Academic Services (IAS), Louisville, KY, USA, September 13, 1989 congratulating him for achievements and inviting him to work in research and teaching in the USA.
- [29] M. Ammar, Performance of Parametric and Reference Pattern Based Features in Static Signature Verification: A Comparative Study, Proceedings of the 10th Int. Conference. on Pattern Recognition, Atlantic City, New Jersey, pp. 646=648, June, 1990.
- [30] M. Ammar, Progress in verification of skillfully simulated signatures, Characters and Handwriting Recognition: Expanding frontiers, P. S. P. Wang, World Scientific, 1991.

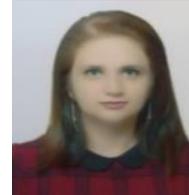
- [31] M. Ammar, M. Aita, B. Younaki, and B. Takwa, A Practical software system for automatic off-line verification of signatures, usable with IBM-PC compatible machines, Seventh Biannual Conference of the International Graphonomics Society, London, Ontario, Canada, Aug. 1995.
- [32] M. Ammar, Certification as Highly qualified Forensic Document Analyst, Justice ministry Decision No. 11/97, Damascus, Syria.
- [33] Maan Ammar, Method and apparatus for verification of signatures, United States Patent: No. 6424728, 07/23/2002, U.S.A.
- [34] Musbah M. Aquel and Maan Ammar, Functions, structures and operation of modern systems for authentication of signatures of bank checks, *Information Technology Journal* (4)1:96-105,2005.
- [35] M. Ammar and M. Aquel, Verification of signatures of bank checks at very low resolutions and noisy images, *Jordan Journal of Applied Science University*, 2005: Vol 7, No. 1, 1-23.
- [36] M. Ammar, Application of Artificial Intelligence and Computer Vision techniques to signatory recognition, *Pakistan Journal of Information and Technology*, 2(1): 44-51, 2003.
- [37] M. Ammar et al., A high efficiency method for automatic signature verification (ASV) in I-C-I environment, *Pakistan Journal of Information and Technology*, 1(2): 160-172, 2002.
- [38] M. Ammar, *Intelligent Signature Verification and Analysis*, Lambert Academic Publishing, Germany, 2011.
- [39] Ammar M., Raising the Performance of Automatic Signature Verification Over that Obtainable by Using the Best Feature Set, *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, Vol. 25, No. 2, 2011, (38 pages).
- [40] Ammar, M., Watanabe, T., Fukumura, T., A New Decision Making Approach for Improving the Performance of Automatic Signature Verification Using Multi-Sets of Features, *International Conference on Frontiers in Handwriting Recognition (Kolkata, India, October 2010)*, pp. 323-328, 2010.
- [41] JA Bullinaria, *Expert systems*, IAI, 2005.
- [42] M. Ammar, *Medical image processing and display (in Arabic)*, p. 464, Damascus university Press, 2013.
- [43] M. Ammar, *Digital Image processing*, R. Gonzalez and P. Wintz, (**Translation** from English to Arabic, 728 pages), The Arabic Center for Arabization, Translation, Authorization and Publishing, Damascus, Syria, 1992.
- [44] M. Ammar, *Biomedical Image Processing*, Damascus University press, Damascus, Syria, 1993 (495 pages in Arabic).
- [45] M. Ammar, *Biomedical Image Display Systems*, Damascus University press, Damascus, Syria, 1994 (384 pages in Arabic).

AUTHORS

Maan Ammar Ph. D. in Information Engineering, Nagoya University, Japan, 1989, Professor at Al Andalus University for medical sciences, Biomedical engineering since 2014, Full professor at Applied Sciences University, Amman Jordan 2003, US patent of a commercial system serving hundreds of US banks since 2002 "Method and apparatus for verification of signatures", United States Patent: No. 6424728 , 07/23/2002, U.S.A. Published tens of papers in image processing and pattern recognition fields. Served as Head of biomedical engineering department–Damascus University for 8 years.



Khuzama Ammar Master Degree in laboratorial analysis. Head of bacteria laboratory at Damascus University educational hospital, Damascus.



Kenan Mansour MD Obstetrics and Gynaecology Specialist, Tishreen Hospital, 2011. Medical Manager of Al Andalus University Hospital, 2019 to present. Medical Education Master student, Syrian Virtual University, 2020. Interested in medical image analysis and diagnosis.



Waad Ammar MD General Surgery Specialist, Tishreen Hospital, 2020. At present, working at Al Andalus University Hospital. Medical Education Master student, Syrian Virtual University, 2020. Interested in medical image analysis and diagnosis.



PHASE DIFFERENCE BASED DOPPLER DISAMBIGUATION METHOD FOR TDM- MIMOFMCW RADARS

Qingshan Shen and Qingbo Wang

College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics, Nanjing, China

ABSTRACT

State-of-the-art automotive radar sensors use a Multiple-Input Multiple-Output (MIMO) approach to obtain a better angular resolution. Time-Division Multiplexing (TDM) scheme is commonly applied to realize the orthogonality in time at the transmitter. Apart from its simplicity in implementation, TDM scheme has the drawback of a reduced maximum unambiguous Doppler proportional to the number of transmitters. In this paper, a phase difference based Doppler disambiguation method is proposed to regain the maximum unambiguous Doppler which is equivalent to only one transmitter. This method works well when the number of transmitters is large. The proposed method is demonstrated with simulation and measurement data.

KEYWORDS

Doppler disambiguation, TDM, MIMO, FMCW, Phase difference.

1. INTRODUCTION

Current-generation automotive radar requires high resolution in the aspect of range, velocity and angle of azimuth. For high-resolution estimation of arrival angle, a wide aperture is indispensable. Frequency-Modulated Continuous Wave (FMCW) radars are commonly used in cars for Advanced Driver Assistant Systems (ADAS) with Multiple-Input Multiple-Output (MIMO) technology to meet these requirements. MIMO radar systems consist of multiple transmitters and multiple receivers, offering a large number of virtual antenna elements and relatively high angular resolution [1], [2]. The signals emanating from multiple transmitters need to be orthogonal and this is mainly implemented with the following approaches: Time-Division Multiplexing (TDM), Frequency-Division Multiplexing (FDM), Code-Division Multiplexing (CDM) and Doppler-Division Multiplexing (DDM). TDM-MIMO is the most intuitive and simple way as each transmitter transmits its own waveform alternatively. Ideal orthogonality can be obtained because there is no overlap between any two transmissions [3], [4].

However, TDM-MIMO scheme results in a reduction in the maximum unambiguous velocity that can be measured by the radar. In order to measure velocity, an FMCW radar transmits multiple chirps separated by time interval T_c . Phase difference induced by this interval can be used to estimate the velocity of targets. TDM-MIMO FMCW radars with N_{TX} antennae enlarge this time interval to $N_{TX}T_c$. The maximum unambiguous radial velocity can be given as

$$V_{max} = \pm \frac{\lambda}{4N_{TX}T_c}, \quad (1)$$

where λ is the wavelength of the radar. Apparently, the maximum unambiguous velocity is reduced by N_{TX} which is a significant drawback if the number of antennae is large.

To regain the true velocity, several disambiguation techniques have been used previously. The Chinese Remainder Theorem (CRT) can be applied for the disambiguation on the foundation of several subsequent measurements with different time intervals [5], [6]. However, when the number of transmitters increases, the CRT algorithm requires more complex time interval configurations and has very limited velocity disambiguation capability. The DBSCAN clustering algorithm can be applied to velocity disambiguation in medium PRF radar and achieve more robustness [7]. Previously neglected high-order phase terms in the received FMCW radar echo were utilized for extension of maximum unambiguous velocity in [8] and a space-time adaptive processing approach was used for Doppler ambiguity in [9]. These methods are somewhat difficult to implement in practice. Recently, Hypothetical Phase Compensation (HPC) technique has been used frequently. This method compensates the velocity induced by phase shift and then selects the correct hypothesis by comparing the peaks of angle FFT results [10], [11]. However, as the number of transmitters increases, the complexity of the calculation increases and the accuracy decreases.

This paper utilizes the phase difference to solve velocity ambiguity by making full use of the phase change information of multiple transmitters and receivers in TDM-MIMO radars. This method can obtain the same maximum unambiguous velocity as with the use of a single transmitter. In this process, no extra hardware costs will be needed and the requirements of calculation will be very simple. The proposed method is validated with simulations and measurement data collected with a 77GHz FMCW cascaded radar. The remainder of the paper is arranged as follows. In Section 2, we analyse the phase difference used for Doppler disambiguation. In Section 3, we analyse the case of Doppler ambiguity and derive a formula for disambiguation based on phase difference. We validate the method through simulation and measurement data in Section 4 and we conclude in Section 5.

2. PHASE DIFFERENCE ANALYSIS

MIMO radar consists of multiple transmitters (TX) and multiple receivers (RX), forming a large virtual array. We can obtain the virtual array signal by performing a two-dimensional FFT (2D-FFT) processing on each TX-RX pair. The range-FFT resolves objects in range and produces a series of bins. A Doppler-FFT is then performed for each range-bin across chirps and thus a signal at a specific range-Doppler bin indicates an object at that range and velocity [5].

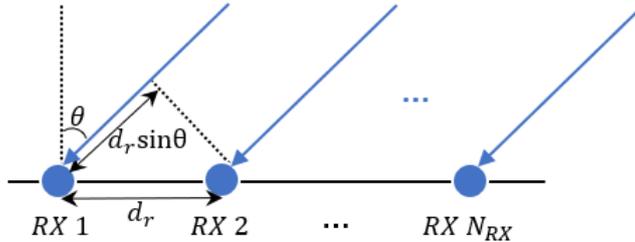


Figure 1. RX antenna array.

Consider an object moving away from radar with a relative velocity v , and denote the angle of arrival by θ . Figure 1 shows the RX antenna array. Two adjacent RX antennae are separated by a distance d_r , the signal from the object must travel an additional distance of $d_r \sin\theta$ to reach the next RX antenna. This distance corresponds to a phase difference between the signals received at two adjacent RX antennae, and the phase difference can be given as

$$\phi_r = \frac{2\pi d_r \sin\theta}{\lambda}. \quad (2)$$

As for TX antennae, the same phase difference can be derived as

$$\phi_{t_azi} = \frac{2\pi d_t \sin\theta}{\lambda}, \quad (3)$$

where d_t is the distance between two adjacent TX antennae. Since TX antennae transmit alternately in a TDM-MIMO radar, an additional phase difference caused by the velocity of object will be introduced. As the time interval between two adjacent TX antennae is T_c , this part of phase difference can be derived as

$$\phi_{t_v} = \frac{4\pi v T_c}{\lambda}. \quad (4)$$

The actual phase difference between two adjacent TX antennae can be expressed as

$$\phi_t = \phi_{t_azi} + \phi_{t_v} = \frac{2\pi d_t \sin\theta}{\lambda} + \phi_{t_v}. \quad (5)$$

Combine (2) and (5), the effect of the angle of arrival can be eliminated and thus we can obtain

$$\phi_{t_v} = \phi_t - \frac{d_t}{d_r} \phi_r. \quad (6)$$

3. DOPPLER DISAMBIGUATION

The relative velocity of the object can be estimated from the Doppler-FFT and denote this velocity by v_{det} . This value can be converted into a phase change

$$\phi_{det} = \frac{4\pi v_{det} N_{TX} T_c}{\lambda}, \quad (7)$$

which corresponds to the phase difference of the same TX antenna between two adjacent chirps. Figure 2 illustrates this phase difference for a positive velocity and a negative velocity.

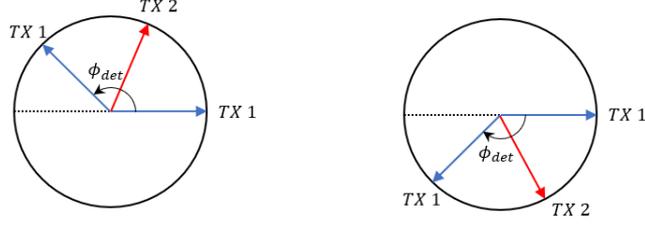


Figure 2. Phase change with a positive velocity (left) and a negative velocity (right).

Doppler ambiguity occurs when the actual phase change exceeds $\pm\pi$. Figure 3 illustrates the actual phase change in the case of two TX antennae and this value can be derived as

$$\phi_{true} = \phi_{det} + 2n\pi, n \in [1 - N_{TX}, N_{TX} - 1] \quad (8)$$

for N_{TX} TX antennae. And this true phase difference can also be obtained as

$$\phi_{true} = N_{TX}\phi_{t.v}. \quad (9)$$

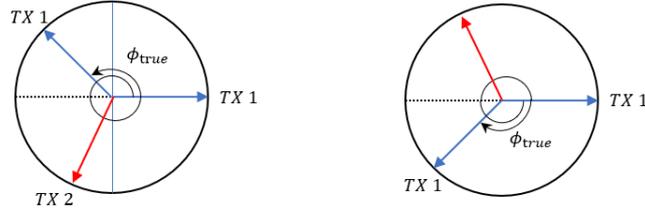


Figure 3. Actual phase change for the situation of two TX antennae, positive velocity (left) and negative velocity (right).

With the combination of (8) and (9), the number of rotations of the phase change can be calculated as follows

$$n = \frac{1}{2\pi} \left[N_{TX} \left(\phi_t - \frac{d_t}{d_r} \phi_r \right) - \phi_{det} \right]. \quad (10)$$

Once we get the value of n , the true velocity can be obtained as

$$v = v_{det} + 2\text{round}(n)v_{max}. \quad (11)$$

4. RESULTS

4.1. Simulation

The data used to validate the method are based on synthetic data generated using the Matlab Radar Toolbox. Parameters of waveform and the corresponding detection capabilities used in this part are shown in Table. 1.

Table 1. Parameters of waveform

Center Frequency (GHz)	77
Wavelength (mm)	3.9
Bandwidth (MHz)	750
Chirp Time (us)	42.67
Number of ADC samples	256
Number of chirps per TX	64
Maximum Velocity (m/s)	1.9

We employ an antenna array of 12 TX antennae and 8 RX antennae in order to verify the availability of this method with a large number of TX antennae. Figure 4 shows the velocity results enlarged by this method. The object directly in front of the radar moves at the velocity varying from -24 m/s to 24 m/s in step of 0.2 m/s. As can be seen from Figure 5, the value of n is very close to the desired result because a large amount of data is provided for estimating.

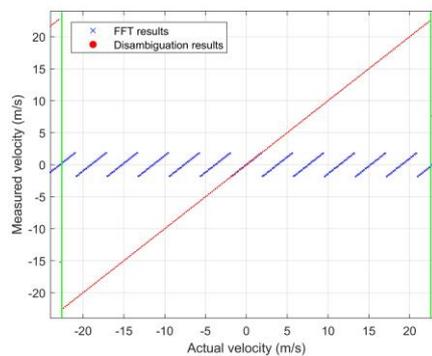


Figure 4. Velocity disambiguation results for an object moving from -24 m/s to 24 m/s, the green lines represent the extended maximum measurable velocity.

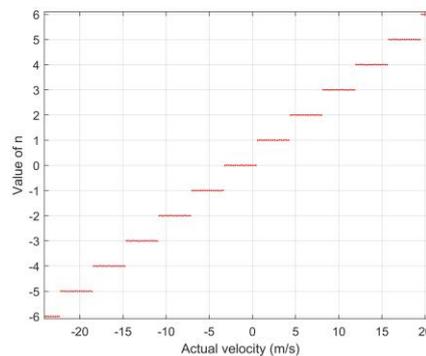


Figure 5. Estimated results of n in simulation.

For situations where the target is not directly in front of the radar, this method can also be available. Figure 6 illustrates the phase information for an object with a velocity of 10 m/s and an angle of 5 degrees to the radar. The RX antennae record information on the azimuth of the object, while the TX antennae record both azimuth and velocity information. Consider an object moving directly ahead at a velocity of 10 m/s, maintaining a certain angle θ to the radar. The radial velocity of the object can be obtained by multiplying by $\cos\theta$. Sweep the angle from -80 degrees

to 80 degrees. The unambiguous velocity results are shown in Figure 7 and the velocity is always retrieved correctly.

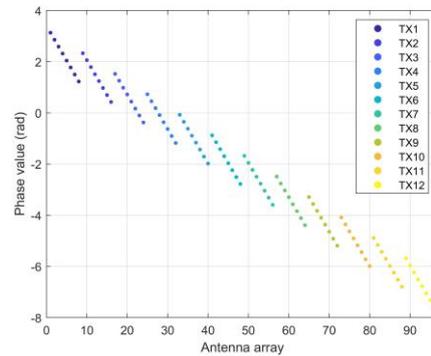


Figure 6. Phase of an object with velocity of 10 m/s and angle of 5 degrees.

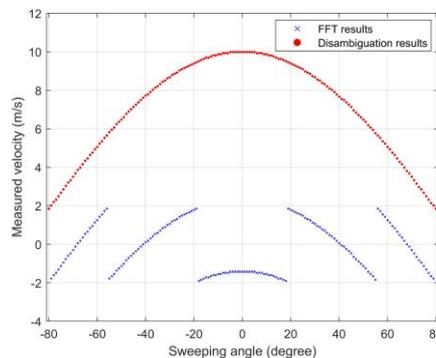


Figure 7. Velocity disambiguation results for sweeping angle.

4.2. Measurement Data

The real data is collected with a commercial 77 GHz cascaded radar from Texas Instruments. This module implements a four-device cascaded array of AWR2243 devices and enables support for up to 12 TX and 16 RX antenna elements. It is worth noting that 3 TX antennae of the cascaded radar are used to measure pitch angle. We can only perform velocity disambiguation by using 9 TX antennae in the horizontal direction. Although the maximum velocity is reduced by a factor of 12, it can still be fully recovered with this method.

The environment for data collection is shown in Figure 8, with many houses and trees on either side of the road, which may cause interference. We collected 50 frames of data with the cascaded radar operating in TDM-MIMO mode. During these frames, the car in front of the radar accelerates from 0 m/s to 8 m/s. As can be seen from Figure 9, all velocity values have been successfully resolved. Figure 10 indicates that the value of n always falls around the correct integer and fluctuations for the reason of noise and clutter. From measured data it can be concluded that this method performs well in real data.



Figure 8. Environment for data collection.

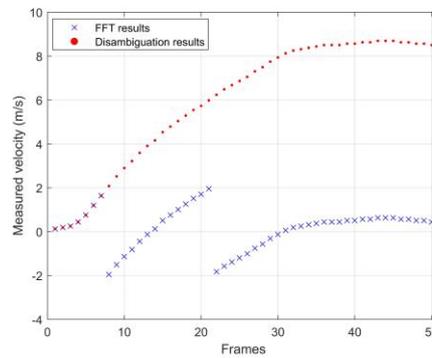
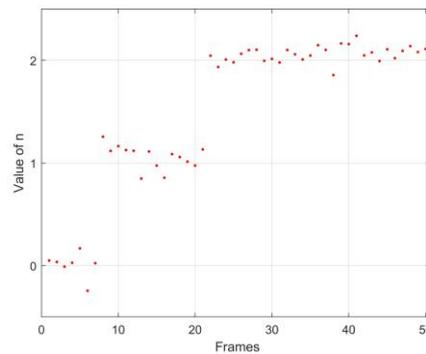


Figure 9. Velocity disambiguation results for a car accelerates from 0 m/s to 8 m/s.

Figure 10. Estimated results of n in measurement data.

5. CONCLUSIONS

This paper proposes a Doppler disambiguation method based on phase difference for TDM-MIMO FMCW radars. This method relies on the measurement of the phase and requires a considerable reliability of the phase estimate. The phase difference provided by RX antennae can be used for compensating the angle of arrival in TX antennae. As a result, we can correctly obtain the unambiguous velocity from different TX antennae. This method makes full use of the data from multiple antennae and does not require any additional conditions such as changing chirp

times or building overlapping elements in the virtual aperture. This approach is particularly effective when the number of antennae is large. Simulation with Matlab Radar Toolbox and measurement results with cascaded radar are presented to validate this method.

ACKNOWLEDGEMENTS

This work was funded by Special Innovation Project for National Defense 19-163-11-ZT-002-002-02. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

- [1] Sun, S., Petropulu, A. P., & Poor, H. V. (2020). MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges. *IEEE Signal Processing Magazine*, 37(4), 98-117.
- [2] Roos, F., Bechter, J., Knill, C., Schweizer, B., & Waldschmidt, C. (2019). Radar sensors for autonomous driving: Modulation schemes and interference mitigation. *IEEE Microwave Magazine*, 20(9), 58-72.
- [3] Sun, H., Brigui, F., & Lesturgie, M. (2014, October). Analysis and comparison of MIMO radar waveforms. In *2014 International Radar Conference* (pp. 1-6). IEEE.
- [4] Rao, S., Subburaj, K., Wang, D., & Ahmad, A. (2020). *U.S. Patent No. 10,627,483*. Washington, DC: U.S. Patent and Trademark Office.
- [5] Zhen-xing, H., & Zheng, W. (1987, April). Range ambiguity resolution in multiple PRF pulse Doppler radars. In *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 12, pp. 1786-1789). IEEE.
- [6] Kronauge, M., Schroeder, C., & Rohling, H. (2010, June). Radar target detection and Doppler ambiguity resolution. In *11-th International Radar Symposium* (pp. 1-4). IEEE.
- [7] Tuinstra, T. R. (2016, July). Range and velocity disambiguation in medium PRF radar with the DBSCAN clustering algorithm. In *2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS)* (pp. 396-400). IEEE.
- [8] Dikshtein, M., Longman, O., Villeval, S., & Bilik, I. (2021). Automotive Radar Maximum Unambiguous Velocity Extension via High-Order Phase Components. *IEEE Transactions on Aerospace and Electronic Systems*.
- [9] Wang, G., & Mishra, K. V. (2020, September). Stap in automotive mimo radar with transmitter scheduling. In *2020 IEEE Radar Conference (RadarConf20)* (pp. 1-6). IEEE.
- [10] GRoos, F., Bechter, J., Appenrodt, N., Dickmann, J., & Waldschmidt, C. (2018, April). Enhancement of Doppler unambiguity for chirp-sequence modulated TDM-MIMO radars. In *2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)* (pp. 1-4). IEEE.
- [11] Liu, C., Gonzalez, H. A., Vogginger, B., & Mayr, C. G. (2021, January). Phase-based Doppler Disambiguation in TDM and BPM MIMO FMCW Radars. In *2021 IEEE Radio and Wireless Symposium (RWS)* (pp. 87-90). IEEE.

AUTHORS

Qingshan Shen was born in Anhui, China, in 1997. He received the B.S. degree in engineering mechanics from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2019. He is currently pursuing the Master's degree in Nanjing University of Aeronautics and Astronautics, Nanjing, China. Currently, the major research focus is on signal processing for high resolution millimetre wave radar.



Qingbo Wang was born in Jiangxi, China, in 1997. He is currently pursuing the Master's degree in Nanjing University of Aeronautics and Astronautics, Nanjing, China. His main research interests are in the estimation of acceleration parameters for millimetre wave radar.



© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

TOWARDS MODI SCRIPT PRESERVATION: TOOLS FOR DIGITIZATION

Kishor Patil, Neha Gupta, Damodar M and Ajai Kumar

Centre for Development of Advanced Computing (C-DAC), Pune, India

ABSTRACT

Modi (मोडई, modī) is a heritage script belonging to Brahmi family, which is used mainly for writing Marathi, an Indo-Aryan language spoken in western and central India, mostly in the state of Maharashtra. "Modi-manuscript "written from the past, reveals the history of the Maratha Empire from its inception under Chhatrapati Shivaji Maharaj; to the creation of movable metal type when Modi was slowly relegated to an inferior position, unfolds perspectives and reflects the social, political and cultural sense of his time." Today it is very important for historians, researchers and students to understand this script and use it for historical heritage. Other regional languages such as Hindi, Gujarati, Kannada, Konkani and Telugu were also using Modi. This paper presents our contribution in helping the community for preserving the script, by way of using various tools, which will facilitate the collection, analysis, and digitization of the Modi script.

KEYWORDS

Language preservation, language development, Modi, Language Analysis, Heritage script.

1. INTRODUCTION

The origin of Modi script is a debatable issue. According to certain historians, Modi can be dated back to the Maurya Dynasty (322–185 BC) and hence the name: 'Modi'. However, the most credible account of the origin of Modi is that Hemadripant, is credited with the invention of the Modi script and the date assigned to its 'birth' is the year 1260.

The invention of Modi can be termed as an act of genius. Modi was invented as a cursive 'shorthand' or speed writing to note down the royal edicts. Traditional Devanagari was found to be excessively time-consuming since each character required as many as 3 to 5 strokes and the lifting of the hand, each time the stroke was completed. Modi got round this obstacle by 'bending' the letters thereby doing away with the need of lifting the hand. This invention thus allowed for a continuous writing which could be used by court scribes to note the edicts. As an example: the handwritten Marathi letter 'has seven 'hand-lifts' whereas' in Modi because of continuous flow requires not a single hand lift. Termed as 'Lapetdar' or Cursive, Modi became extremely popular and rivaled the 'Shikasta' script of Persian. The introduction and history about the Modi script has been covered by the [1][2] and [3] references.

2. LITERATURE SURVEY

Modi is a heritage script and not much of the content is available on the internet. However, the proposal by Anshuman Pandey, "Proposal to Encode the Modi Script in ISO/IEC 10646" covers

the details about the Modi Script [5]. The proposal covered the background and writing system of the Modi script which is a great starting point to understand the script. The proposal also covers characters and character combinations in detail, which is very helpful to understand the orthography of the script. The article covered all the aspects of the Modi script including conjuncts formations, head strokes, word and section boundaries and collating order. Comparison of Modi script and Devanagari has also been covered in the article. With the Author's efforts, Modi is now a part of the Unicode Standard [6]

11600

Modi

1165F

	1160	1161	1162	1163	1164	1165
0	ॐ 11600	ग 11610	घ 11620	ॠ 11630	ॡ 11640	ॢ 11650
1	ॣ 11601	घ 11611	॥ 11621	ी 11631	 11641	१ 11651
2	॥ 11602	ड 11612	ॢ 11622	ी 11632	 11642	२ 11652
3	॥ 11603	उ 11613	ॣ 11623	ु 11633	० 11643	३ 11653
4	॥ 11604	छ 11614	घ 11624	ू 11634	J 11644	४ 11654
5	॥ 11605	॥ 11615	ॣ 11625	ृ 11635		५ 11655
6	॥ 11606	॥ 11616	ॣ 11626	ृ 11636		६ 11656
7	॥ 11607	॥ 11617	ॣ 11627	ृ 11637		७ 11657
8	॥ 11608	॥ 11618	ॣ 11628	ृ 11638		८ 11658
9	॥ 11609	॥ 11619	ॣ 11629	ृ 11639		९ 11659
A	॥ 1160A	॥ 1161A	ॣ 1162A	ृ 1163A		
B	॥ 1160B	॥ 1161B	ॣ 1162B	ृ 1163B		
C	॥ 1160C	॥ 1161C	ॣ 1162C	ृ 1163C		
D	॥ 1160D	॥ 1161D	ॣ 1162D	ृ 1163D		
E	॥ 1160E	॥ 1161E	ॣ 1162E	ृ 1163E		
F	॥ 1160F	॥ 1161F	ॣ 1162F	ृ 1163F		

The Unicode Standard 14.0, Copyright © 1991-2021 Unicode, Inc. All rights reserved.

Figure 1. Modi Unicode code chart

3. THE EVOLUTION OF MODI

Modi script remained in practice from the 13th century to the first half of 20th century. During the seven odd centuries, Modi calligraphy underwent several transitions. The official correspondence of Maharashtra as well as the regions like Madras, Mysore, Bundelkhand, Gujarat, and Rajasthan was written in Modi for a long time. Transitions in Modi calligraphy fall in four periods Bahamani, Shivkalin, Shahukalin and Anglakalin. In each of these periods the 'style' and 'lapeti' of Modi calligraphy underwent a considerable change.

3.1. Bahamani kalin Modi

Modi script in the Bahamani period was impacted by Perso-Arabic script. The letter of Shahaji Maharaj (18 March 1594 - 23 January 1664) shows script structure of Bahamani era (Figure 2).

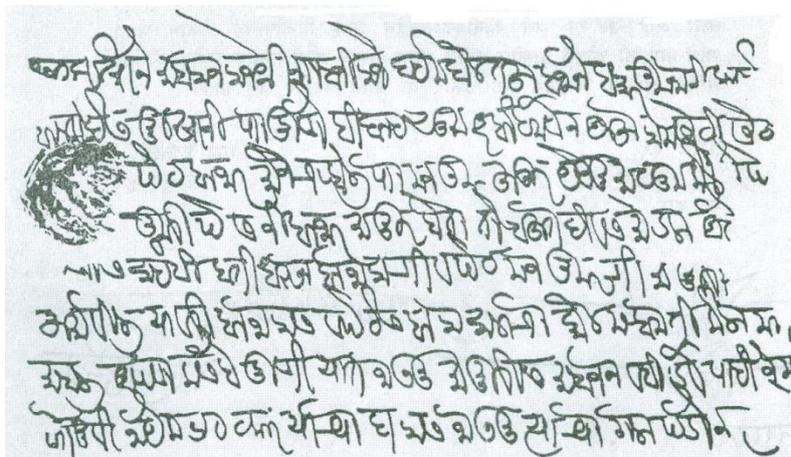


Figure 2. Letter of Bahamani kalin Modi

3.2. Shivkalin Modi

Modi script in the period of Chhatrapati Shivaji Maharaj (19 February 1630 - 3 April 1680) period was also impacted by Perso- Arabic script. The letter of Chhatrapati Sambhaji Maharaj shows the beauty of Shivkalin era (Figure 3, Figure 4).

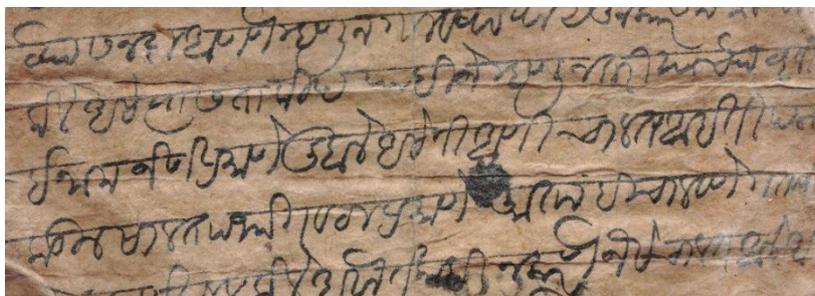


Figure 3. Letter of Shivkalin Modi

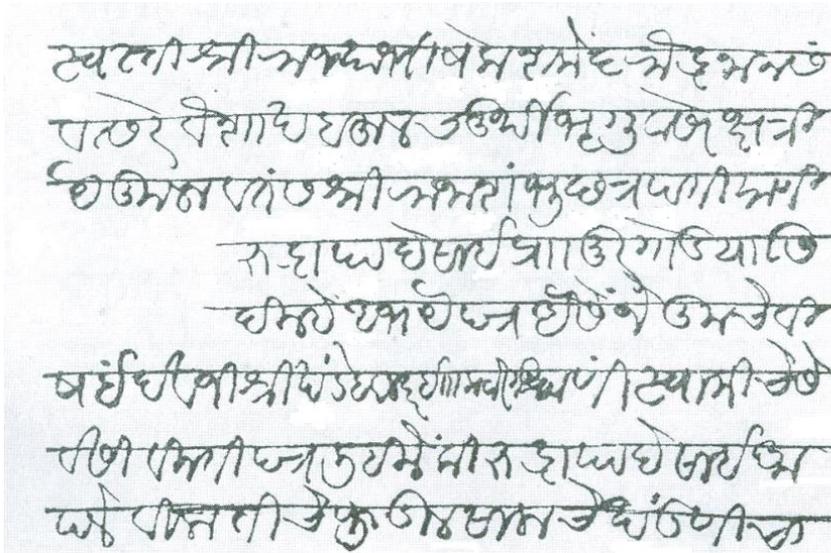


Figure 4. Letter of Shivkalin Modi

3.3. Shahukalin Modi

Modi script in the period of Chhatrapati Shahu Maharaj (18 May 1682 - 15 December 1749) and Peshwa's is more cursive. The curves are more marked and elongated. The curves in fact gave rise to various styles such as Mahadajipanti, Biwalkari and Ranadi (Figure 5).

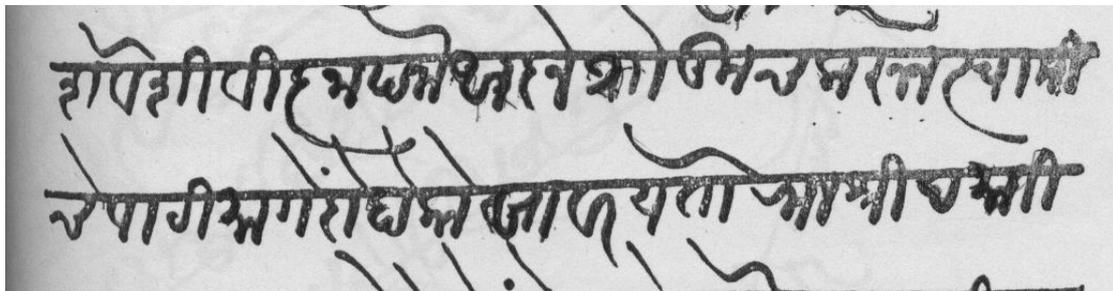


Figure 5. Letter of Shahukalin Modi

3.4. Anglakalin Modi

Under the first half of the British period (1818-1947) English and Modi co-existed together. Most of the correspondence in the Deccan was in Modi and advertisements in English newspapers were also in Modi. However with the innovation of Moving Metal-type and the spread of English language the downfall of Modi script became certain. The Modi of the British period is influenced by the use of the pen to write letters. Thick-thin variants show their presence for the first time as in the following specimen of a letter of General Grant Duff (Figure 6).



Figure 6. Letter of Anglakalin Modi

4. MODI WRITING SYSTEM DETAILS

Although Modi is inherited the model as Devanagari, it differs considerably from the Devanagari in terms of letter forming, rendering behaviours, and orthography. Modi was invented as a cursive “shorthand” or speed writing to note down the royal edicts. The traditional Devanagari turned out to be too laborious since each character required up to 3-5 strokes and the raising of the hand, each time the stroke was completed. Modi got found this obstacle by “bending” the letters thereby doing away with the need of lifting the hand. This invention thus allowed for a continuous writing which could be used by court scribes to note the edicts (Figure 7, Figure 8, Figure 9, Figure 10, Figure 11, Figure 12). However, this 'speed-writing' led to certain modifications of which the most notable features are as under:

- The total absence of short and long vowel forms as well as vowel modifier forms
- Use of certain specific markers used in Modi as prefixes for numerical notation
- Practically no derived ligature forms: conjuncts being marked either by use of the virama or by use of half characters

More details about the Modi alphabets is covered in the link [4]



Figure 7. Modi Vowels

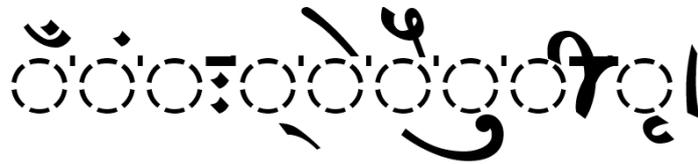


Figure 8. Modi Vowel Signs



Figure 9. Modi Numerals

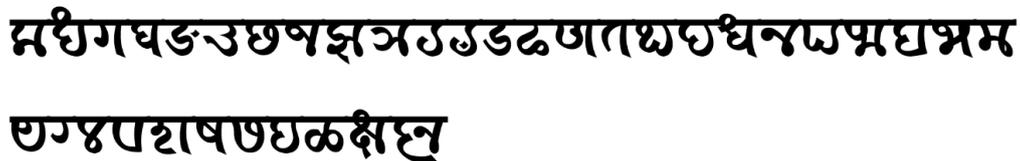


Figure 10. Modi Consonants

The shapes of some consonants, vowels, and vowel signs are similar. The actual differences are visible when characters are part of consonant-vowel combinations / consonant conjuncts.

कधगाघडाउछभझत्रालजडढाजाताधघधान्नपापूधन्न
मनामप्रशाषाजघकाकैधन

Figure 11. Modi Consonants-aa matra combinations

तान्नपशुउशुजाधुधुन्यषुषुन्याधुषु

Figure 12. Modi Conjunctions

5. COMPARISON BETWEEN MODI AND OTHER SCRIPTS

This section compares Modi and Other scripts. Modi and various other scripts, as shown below, have some similarities, as they are either variant or influences from Devanagari. The below table (figure 13, figure 14) shows the consonants comparison between Modi, Mahajani [7], Landa [8], Kaithi [9], and Devanagari [10].

	Modi	Mahajani	Landa	Kaithi	Devanagari
KA	क	क	क	क	क
KHA	ख	ख	ख	ख	ख
GA	ग	ग	ग	ग	ग
GHA	घ	घ	घ	घ	घ
NGA	ङ	—	ङ	ङ	ङ
CA	च	च	च	च	च
CHA	छ	छ	छ	छ	छ
JA	ज	ज	ज	ज	ज
JHA	झ	झ	झ	झ	झ
NYA	ञ	ञ	ञ	ञ	ञ

Figure 13. Consonants comparison (ka-nya)

	Modi	Mahajani	Landa	Kaithi	Devanagari
TTA	उ	८	८	८	ट
TTHA	उ	०	०	०	ठ
DDA	ड	१	३	७	ड
DDHA	ढ	७	८	७	ढ
NNA	ण	२॥	८	३॥	ण
TA	त	२	३	१	त
THA	थ	३	४	थ	थ
DA	द	६	४	६	द
DHA	ध	२	५	५	ध
NA	न	७	८	७	न
PA	प	५	५	५	प
PHA	फ	३	७	३	फ
BA	ब	५	७	५	ब
BHA	भ	९	३	७	भ
MA	म	५	४	५	म
YA	य	—	७	५	य
RA	र	०	१	१	र
LA	ल	३	७	७	ल
VA	व	५	६	५	व
SHA	श	—	—	श	श
SA	स	७	७	५	स
HA	ह	॥	५	६	ह

Figure 14. Consonants comparison (tta-ha)

6. COLLECTION OF TOOLS

This section explains existing work that are related to the Modi Puratan Dastavej Jatan Pranali - Digital Annotation & Archiving System

6.1. MODI-SHAHU (मोडी-शाहू) Font

MODI-SHAHU font is specially designed font as per the letter forms, rendering behaviour and orthography. This font fulfils all the basic requirements such as Structure and form of the character for better digitization (Figure 15).

— श्री —

नाभमन्लनाभश्रीघनीनापतिजंतिप्रधानमंठीथद्वमेरु
 औषीनेषुल्लरेनेगीउेघँषुषु००च्लुषेघप
 षेगीचसभपमषषेष्टेचनानेगीप्रषेपल्लरुमिनथ
 चरुमीमरीतीशृीभिमीथजणामडिषथुमषदांषुता
 मीचमसन्दिभमीमन्घेताषुषुपापेतोचल्लतीउतागा
 दान्नगेतेगोशीमरुषेपजेजेउेघंरिथंभृंनभमीठ
 कडेघरांदिच्लपमउाठीस्ताथेह्लषुषुषुषुषु
 प्रथीषेपघरीमंजीरुस्मीषमसन्दाघंदिप्रो।मीठषेपी
 षेतीषेपनीठीस्ताषुथंमरापेघेतिमभमीकतिपरीह्ल
 षुतामीचमसन्चरुषेकडेघरांदिप्रमणेमीठषेपणे
 यथीह्लथुषु

Figure 15. Modi-Shahu Font

6.2. MODI Typing Tool

Modi Typing tool is developed for inputting the content /text in Modi script. The keyboard is designed using the INSCRIPT principle. The keyboard is based on Unicode; hence, the documents thus created will be easily be viewed properly on any Unicode enabled operating systems such as Windows 10. On-screen keyboard for Modi Script is also provided in this tool to make typing more easily. The tool uses the font "MODI-SHAHU" for rendering the character on the Keyboard (Figure 17).



Figure 16. Modi Typing Tool UI



Figure 17. Modi Keyboard

6.3. MODI - DEVANAGARI Converter

The Modi-Devanagari converter converts data from Modi script to Devanagari script. The converter is designed based on research and study of Modi and Devanagari scripts written in Marathi language (Figure 18).

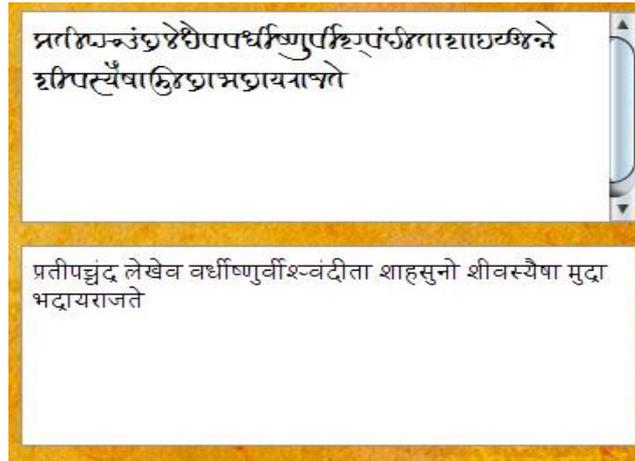


Figure 18. Modi-Devanagari Converter

6.4. Modi Puratan Dastavej Jatan Pranali - Digital Annotation & Archiving System

The "Modi Puratan Dastavej Jatan Pranali (मोडी पुरातन दस्तावेज जतन प्रणाली) Annotation and Digital Archiving System" is a web application system for the digital preservation of cultural heritage resources and manuscripts for Modi. The system takes historical Modi documents available as images and allows the user to annotate and type in the Modi text. A virtual keyboard is provided for easy typing of Modi script. A transliteration system is also provided to convert Modi text to Devanagari. System uses the specially designed font MODI-SHAHU for proper display of content in the Modi Script. (Figure 19, Figure 20, Figure 21, Figure 22).



Figure 19. Modi Puratan Dastavej Jatan Pranali - Start Screen

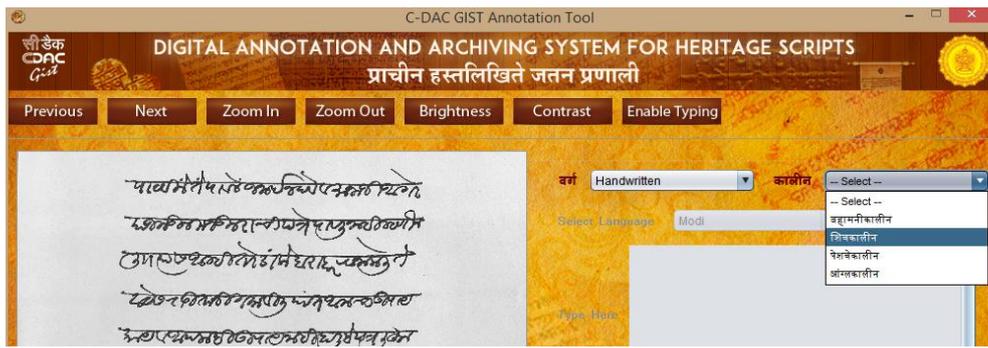


Figure 20. Modi Puratan Dastavej Jatan Pranali - Selection of Era

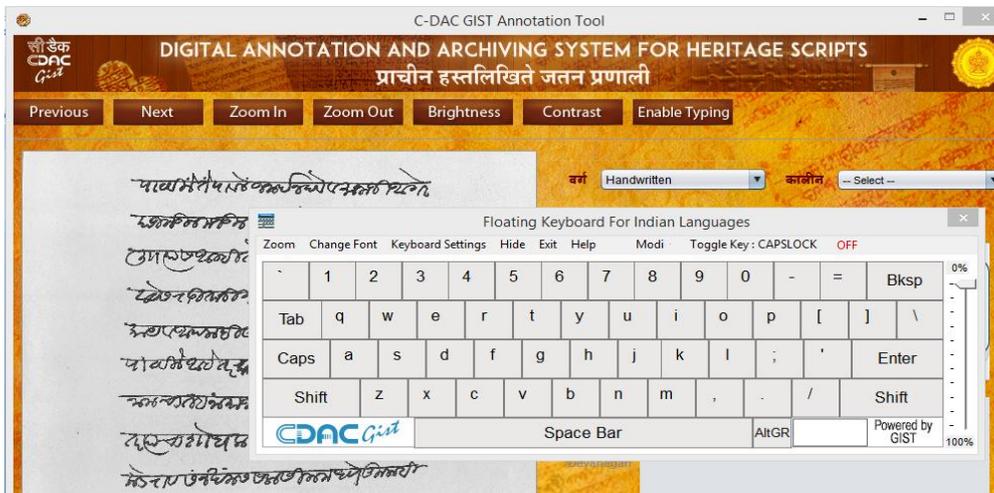


Figure 21. Modi Puratan Dastavej Jatan Pranali - Keyboard

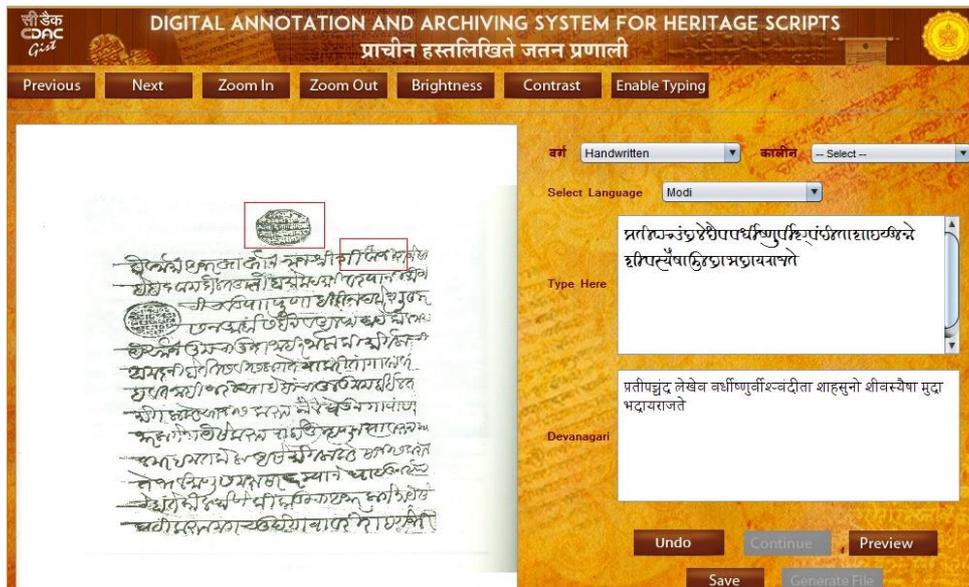


Figure 22. Modi Puratan Dastavej Jatan Pranali - Annotation and converter screen

7. CONCLUSION

The development of "Modi Puratan Dastavej Jatan Pranali" is a functional tool for preserving the script from extinction. It also facilitates various tools and technologies for the community user to better understand the script and digitize it. "Modi Puratan Dastavej Jatan Pranali" is a collection of various tools by which data available in the form of images will be easily inputted, displayed, converted, and stored in a database. However if further applications such as OCR, the search engine could be developed and integrated with it, it can increase the usability of the existing systems, which is a future proposition.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Information Technology (DIT), Maharashtra for funding this project.

REFERENCES

- [1] "Modi script," https://en.wikipedia.org/wiki/modi_script.
- [2] "History of modi lipi," <http://www.modilipi.in/2011/02/modi-script-of-maharashtra-script-which.html>.
- [3] "The origin and development of indian writing system," <https://narendranath.webs.com/>.
- [4] "Modi alphabet," <https://omniglot.com/writing/modi.htm>.
- [5] "Proposal to Encode the Modi Script in ISO/IEC 10646", <http://unicode.org/L2/L2011/11212r-n4034-modi.pdf>
- [6] "Modi Unicode Code Chart", <https://www.unicode.org/charts/PDF/U11600.pdf>
- [7] "Mahajani Script", <https://omniglot.com/writing/mahajani.htm>
- [8] "Landa Script", <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3766.pdf>
- [9] "Kaithi Script", <https://omniglot.com/writing/kaithi.htm>
- [10] "Devanagari Script", <https://omniglot.com/writing/devanagari.htm>

AUTHORS

Mr. Kishore Patil has worked as a typographer, designer and visualizer for over 20 years. He has worked in various fields of fine arts, such as advertising, newspapers, agricultural research, medical research and the IT sector. His interest in designing and developing Open Type fonts for all Indian languages and heritage scripts.



Ms. Neha Gupta has more than 14 years of experience in projects related to research and development of local language technology and standardization initiatives. Her specialization includes developing language processing tools/systems viz. Internationalized Domain Names implementation at India Level, Neural Machine Translation (NMT), NER, POS to name a few specifically with AI/ML/DL and rule based methodologies.



Dr. Damodar Magdum has been working in Indian language technology for over 15 years. His specialization in the creation of text to speech (TTS) corpus in Indian language. He has special interest in the proliferation and preservation of heritage language in India.



Dr. Ajai Kumar is Senior Director and Head of the Applied Artificial Intelligence & GIST Group and have more than 20 years of experience working in Natural Language Processing including Machine Translation, Speech Technology and Information Extraction & Retrieval and E-learning systems. His key role is in initiating mission mode consortium projects in the area of Natural Language Processing, language learning, Video Surveillance etc.



© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

TASK-ORIENTED DIALOGUE SYSTEMS: PERFORMANCE VS QUALITY- OPTIMA, A REVIEW

Ryan Fellows^{1*}, Hisham Ihshaish¹, Steve Battle¹,
Ciaran Haines¹, Peter Mayhew^{1,2}, J. Ignacio Deza^{1,3}

¹ Computer Science Research Centre (CSRC),
University of the West of England (UWE), Bristol, United Kingdom

² GE Aviation, Cheltenham, United Kingdom

³ Universidad Atlántida Argentina, Mar del Plata, Argentina

ABSTRACT

Task-oriented dialogue systems (TODS) – designed to assist users to achieve a goal – are continuing to rise in popularity as various industries find ways to effectively harness their capabilities, saving both time and money. However, even state-of-the-art TODS have not yet reached their full potential. TODS typically have a primary design focus on completing the task at hand, so the metric of task-resolution should take priority. Other conversational quality attributes that may point to the success, or otherwise, of the dialogue, are usually ignored. This can harm the interactions between the human and the dialogue system leaving the user dissatisfied or frustrated. This paper explores the role of conversational quality attributes within dialogue systems, looking at if, how, and where they are utilised, and examining their correlation with the performance of the dialogue system.

KEYWORDS

Dialogue Systems, Chatbot, Conversational Agents, AI, Natural Language Processing, Quality Attributes.

1. INTRODUCTION

Dialogue systems, by nature, are typically either chat-oriented or task oriented [1]. Chat-oriented, or conversational, dialogue systems have the objective of relaying contextually appropriate and stimulating responses [2], whereas task-oriented dialogue systems (TODS), or transactional systems, are designed to assist a user in completing their goals. Examples include finding transport times, booking tickets or customer support [3].

Over recent years, the adoption of TODS has surged significantly, as companies recognise their potential in alleviating the resource requirements inherent in human-based dialogue services. A prediction by market research firm Grand View Research estimates that the global chatbot market will reach \$1.23 billion by 2025 [4, 5].

The literature exploring TODS performance generally focuses on benchmarking against human-generated supervised feedback, such as that of task-resolution [6, 7]; a measure that encapsulates the dialogue system's success rate in resolving a task or set of tasks. A direct correlation is assumed between task resolution and the performance of the dialogue system as a whole.

David C. Wyld et al. (Eds): SIPP, NLPCL, BIGML, SOEN, AISC, NCWMC, CCSIT - 2022

pp. 69-87, 2022. CS & IT - CSCP 2022

DOI: 10.5121/csit.2022.121306

Whilst task-resolution is a priority – as the journey of the whole conversation is considered – performance and user experience cannot be disregarded, as they have the potential to hinder adoption of the system, independently of its performance. For this reason, in addition to task-resolution within TODS performance evaluation studies, more in particular compared to other types of dialogue systems, user satisfaction is commonly considered as another performance metric, as an indicator of system efficiency [8, 9] or usability [7].

The user satisfaction metric assumes a relative usability or efficiency for a dialogue system on the basis of how its users are satisfied. These are usually approximated by two approaches: either by means of laboratory experiments, eliciting human judgment on system outputs and behaviour relative to a predefined set of interaction parameters (e.g. number of turns [10], dialogue duration [11]). Or through modelling satisfaction, whereby the aim is to create models that provide ratings of performance similar to those which humans would do. The ratings based on human judgment are then used as target labels to learn an evaluation model based on objectively measurable performance attributes [12].

Comparing the performance of dialogue systems is a non-trivial task. This is due to the wide range of domains in which the systems are deployed, and the criteria they are evaluated against. Interactions are also subjective. What might be an optimal response for one individual, could be completely unsuitable for another, with performance being gauged on that specific individual's communicative preferences.

This paper explores quality attributes that describe different qualities of conversational interactions between a system and the user, besides task outcome. We analyse conversational quality attributes in TODS and explore how they are utilised, and to what effect. To accomplish this, a literature survey is undertaken to examine current considerations to conversational quality attributes used in conjunction with dialogue systems.

Throughout this paper, adherence will be made to a real-world locally collected corpus of interactions between University students and staff and University helpdesk assistants. This dataset consists of 600 email threads and 5697 subsequent emails - which are made up of a sender direction (incoming or outgoing), subject, body and a time stamp. Interactions consist of a range of issues which students and staff are in need of resolving. This GDPR compliant dataset will be referred to as the *ITS helpdesk* dataset throughout this paper.

The rest of the paper is organised as follows: Section 2 explores TODS conversational quality attributes and surveys their application to study and evaluate dialogue systems. This section is broken down into sub-sections consisting of individual quality attributes. Further discussion and conclusions are provided in Section 3.

2. CONVERSATIONAL QUALITY ATTRIBUTES

In a real world, human-to-human, task-oriented interaction, a conversation would likely not be deemed successful if only the task was resolved. If the advisor, in this situation, was friendly, personable and efficient in their manner, the advisee would be significantly more likely to have a positive experience and return in the future. However, if the advisor was rude or did not convey information competently, the advisee would most likely be left frustrated or even angry, leaving with a bad impression. Of course, interactions with a human do not translate perfectly to interactions with machines, yet findings from real world communication can be extrapolated and applied to virtual communication.

In most circumstances, a TODS should elicit a positive user experience while seeking to resolve tasks in the most effective way possible. Accordingly, the evaluation of TODS performance generally seeks to optimise two main qualities: task-resolution and dialogue efficiency.

This section surveys the state-of-the-art developments on conversational quality attributes in the context of TODS, and highlights some of the most prominent attributes addressed in the literature around TODS performance.

2.1. Task Resolution

Task-resolution, or goal completion, is one of the most accessible metrics — and arguably can be the easiest to derive given a well-defined user goal as well as a predefined function to quantify a resolved, unresolved or somewhere between, task — to evaluate the success of a TODS. The main purpose of a TODS is to assist a user with a specific task in an automated fashion. Therefore, the success of a dialogue system in fulfilling information requirements established by user goals is an indicator of a dialogue system's performance.

Practically, task-resolution (or success) is used to test dialogue systems success in providing not only the correct information, but also all user requested information — addressing as such the components for a given user-task: a set of constraints (target information, or information scope) and a set of requests (all required information) [13]. This in fact is consistent with the established understanding in Psychology around the notion of ‘conversation’, that is, it is understood that when individuals engage in conversation, there is a mutual understanding of the goals, roles and behaviours that can be expected from the interaction [14, 15]. Therefore, the ‘performance’ of the dialogue has to be evaluated on the basis of their mutual understanding and expectations.

In its simplest form, however, this metric can be quantified as a Boolean — binary task success (BTS) — value indicating whether a task or set of tasks has been resolved or not. Using this metric, organisations can capture useful statistics over a number of interactions to derive how effective their dialogue system is at solving tasks, in comparison to interactions with human assistance or even other dialogue systems.

One of the more inherent challenges of task-resolution, as a performance metric, is knowing whether the task in question has been resolved. Especially so as the different users may have different goals, or intrinsically multiple goals, and these may even change in response to system behaviour throughout the course of interaction. On top of this, different users may have varying definitions of success, for example, a domain-specific expert user may deem a task resolved with less detailed information acquired compared to a novice user.

Typically, an interaction with a dialogue system will end when a user terminates the conversation, however this doesn't necessarily imply that their goals have been met. Some dialogue systems opt to explicitly elicit ‘task completion’ in some form: “*has your request been resolved?*” or “*is there anything else I can help you with?*”, others attempt to use some form of classifier to infer when a task has been resolved through a machine learning and NLP model (eg. [16, 17]). This requires a structured definition of goals and a mechanism to measure success relative to that goal. In this fashion, much of the work on automating the evaluation of task success has largely focused on the domain-specific TODS. This is usually an easier task as such systems can be highly scripted, and task success can be specifically defined – especially so in traditional dialogue systems, such as the Cambridge Restaurant System [18] and the ELVIS email assistant [19] — where the relevant ontology defines intents, slots and values for each slot of the domain.

However, a structured definition of goals will usually bind dialogue systems to a specific class of goals, constraining their ability to adapt to the diversity and dynamics of goals pertinent in human-human dialogue [20]. To address the shortcomings in adaptability and transferability encountered in single-domain systems, research into domain-aware, or multi-domain, dialogue systems has attracted noticeable attention in recent years [21, 22]. This saw the introduction of the concept of the domain state tracker (DST), which accumulates the input of the turn along with the dialogue history to extract a *belief* state: user goals/intentions expressed during the course of conversation. User intentions are then encoded as a discrete set of dialogue states, i.e., a set of slots and their corresponding values, as shown in e.g., [23, 24]. As a result, the multiple user intentions are subsequently evaluated, whether objectively met or otherwise - Please refer to Figure 1 in [25] for a detailed characterisation of DSTs.

Reinforcement learning systems aim to find the optimal action that an automated agent can take in any given circumstance, by either maximizing a reward function or minimizing a cost function. With a dialogue system as the agent, the given circumstance is the belief state held by the DST, the reward function is linked to task-resolution, and the actions are the system's output slots and values. Dialogue systems will inevitably encounter problems; examples include incorrectly identifying a word, or a user changing their goal. A system could assign confidence levels to the belief states, track multiple belief states, and include a plan to recover the conversational thread after the errors are noticed.

Casting the conversation as a partially observable Markov decision process (POMDP) allows for these uncertainties to be encoded [31]. A POMDP is defined as a tuple $\{S, A, \tau, R, O, Z, \lambda, b_0\}$ where S is a set of states describing the environment; A is a set of actions that may be taken by the agent; τ is the transition probability $P(s' | s, a)$; R defines the expected reward $r(s, a)$; O is a set of verifiable observations the agent can receive about the world; Z defines an observation probability, $P(o' | s', a)$; λ is a geometric discount factor $0 \leq \lambda \leq 1$; and b_0 is an initial belief state $b_0(s)$.

A POMDP dialogue system tracks multiple parallel belief states, selecting actions based on the belief state that is most likely. When misunderstandings occur, the current belief state can be made less likely, allowing the system to move to a new belief state. Because the belief states' probabilities are tracked alongside the expected action rewards and the chance that an action will transition as expected, a POMDP is able to effectively plan how to manage a dialogue. This framework allows a TODS to track multiple possible user goals, to plan error checking of user utterances, and to use context to potentially identify when the dialogue system has misunderstood the user intention. However, converting this potential benefit into practice is not trivial. Such systems are known to require a significant amount of training, as the state - action space can be very large even for single domains, and uncertainty in the task resolution may weaken the agent's learning [26].

In general, task-resolution is commonly quantified as the result of a performance metric in which user satisfaction is maximised. The PARADISE framework [27], which is frequently used as a baseline for task success evaluation throughout literature, values user satisfaction as a weighted linear combination of task success measures side by side with dialogue costs (reported in Sec. 2.5). These measures can be objective, which entail features such as word error rate [28], automatic speech recognition (ASR), word-level confidence score [29], number of errors made by the speech recognizer [30] and time to fire, task completion rate, and accuracy metrics as used in [31], or subjective such as intelligibility of synthesized speech [32] and perception tests [33].

Table 1 breaks down the threads within the ITS helpdesk dataset into the task resolution percentage via topic. In this example, threads have been classified into groups of topics using the

unsupervised topic modelling algorithm of latent Dirichlet allocation (LDA) to provide a baseline example of topic categorisation. Threads have now been contextualised to some degree which can then allow further analysis in conjunction with the objective measures that will follow this Section.

Table 1. Breakdown of Task resolution statuses of ITS Helpdesk threads.

Topic	Topic Keywords	Number of Threads	Task Resolution Percentage
1	Person, Need, Would, Email, Work	89	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
2	Student, Person, Access, Look, Module	19	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
3	Access, File, Document, Try, Help	24	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
4	Order, Laptop, Could, Login, Generic	7	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
5	Person, System, Group, Purchase, User	20	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
6	Screen, Mark, Drive, Room, File	17	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
7	Folder, Course, Number, Upload, Video	21	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
8	Person, Add, Address, Email, Staff	311	Resolved: 82% Unresolved: 14.6% N/A: 3.4%
9	Desktop, Office, Slow, Urgently, Computer	92	Resolved: 82% Unresolved: 14.6% N/A: 3.4%

2.2. Usability and Dialogue Efficiency

Usability attributes, such as user satisfaction, learnability, efficiency, etc, are the foundation of the design of ‘successful’ dialogue systems, as these are ultimately created for the user, and for the user to achieve their intended, and occasionally variable, goal(s). While such attributes should ultimately be the criteria to evaluate a dialogue system, they are well-known to be subjective, and subsequently hard to measure. This is why much literature on evaluating dialogue systems tends to deal with quantifiable performance metrics, like task-resolution rate or elapsed time of the interaction. It has been proposed, however, that an agent's competence in objectively measurable dialogue does not necessarily induce a better user experience, and subsequently a better overall usability [34]. In fact, the different metrics may even prompt contentious interpretations, or simply contradict each other [35].

Although usability ratings are notoriously hard to interpret, especially if the system is not equipped to infer and keep track of user goals, the successful encapsulation of such values can provide insight that explicit metrics struggle to capture. From the study of Malchanau et al, usability experts rated examined questions from a 110 item questionnaire and derived an

evaluation of their agreement of usability concepts. This led to a collection of 8 attributes they saw as key factors: task completion and quality, robustness, learnability, flexibility, likeability, ease of use and usefulness (value) of an application [34]. This questionnaire was used to evaluate a dialogue system designed for training purposes, in which the overall system usability was determined by the quality of agreements reached, by the robustness and flexibility of the interaction, and by the quality of system responses.

Additionally, these different metrics may in fact have an inconsistent statistical interpretation to different designers. In the same way human evaluation will provide different outcomes based on the subjective criteria, the same can be said for metrics of usability which are difficult to consistently quantify [35].

2.3. User Sentiment

Because of the insights sentiment analysis reveals about the more concise bodies of text on social media, the field of sentiment analysis has seen a take-up of use over recent times [36]. Sentiment analysis can be performed on large quantities of tweets and posts from different platforms to assess general opinion about a specific product or topic.

Different applications use a range of machine learning classification algorithms to categorise sentiment scores [37, 38], some use just two classes: positive and negative, while others use an n-point scale, e.g., very good, good, satisfactory, bad, very bad [39]. A review and a comparative study of existing techniques for opinion mining like machine learning and lexicon-based approaches is provided in [40].

Table 2. Main user sentiment studies in dialogue systems reviewed in the literature

Domain	Author	Year	Proposal / Findings
SDS	Schuller [41], Nwe[42]	2003	Emotion recognition in spoken dialogue using phonic features.
SDS and TOSS	Devillers [43]	2003/05	Automatic and 'robust' cues for emotion detection using extra linguistic features, lexical and discourse context.
SDS	TH Bui [44]	2006	'Affective' dialogue model: inferring user's emotional state for an adaptive system's response. Earlier work applied to spoken dialogue systems in.
TODS and SDS	Ferreira [45], Ultes [46]	2013/17	Proposed an expert-based reward shaping approach in dialogue management, and a live user satisfaction estimation model based on 'Interaction Quality', a "less subjective variant of user satisfaction".

DS	Shin [47]	2018	Detecting user sentiment from multimodal channels (acoustic, dialogic and textual) and incorporating the detected sentiment as feedback into adaptive end-to-end DS
DS	Jaques [48]	2019	Deep reinforcement learning model (off-policy batch RL algorithm).
DS	Shin [49]	2019	Happybot: on-policy learning in conjunction with a user-sentiment approximator to improve a seq2seq dialogue model.
DS	Sasha [50]	2020	Applying Reinforced Learning to manage multi-intent conversations with sentiment based immediate rewards

DS: Dialogue Systems, **SDS:** Spoken Dialogue Systems, **TODS:** Task-oriented Dialogue Systems, **TOSS:** Task-oriented Spoken Systems

Early studies on sentiment analysis in the context of dialogue systems explored the inclusion of user sentiment in rule-based systems, towards adaptive spoken dialogue systems [51, 52]. Most of these studies investigated modular-based dialogue systems (conventionally referred to as pipeline models), with predefined rules for systems to adapt to variability in user sentiment. In recent studies, however, much focus has been placed onto sentiment-adaptive end-to-end dialogue systems, particularly due to their adaptability in comparison with modular-based ones [53], which are known to be harder to train, and adapt to new contexts [54].

Studies exploring the conjunction of dialogue systems with sentiment analysis are often motivated by the notion of system *adaptability*, assuming a correlation between adaptability of the systems to user sentiment and their satisfaction. Some recent work emphasises the importance for conversational agents to adapt to different user (personality) types [55, 56]. Attention is paid to studying user sentiment as a variable to guide the design of sentiment-adaptive dialogue systems [57, 58]. A comprehensive list of development milestones on sentiment analysis application to the analysis and evaluation of dialogue systems, as well as on sentiment-adaptive systems is provided in Table 2.

It should be noted, nonetheless, that sentiment analysis methods have not been extensively applied to conversational agents and dialogue systems. One reason for this is the fact sentiment analysis performs more effectively when pre-trained on a domain specific dataset, and would not often generalise to open domains of discourse inherent in many dialogue systems. One example is the well-known shortcomings when generalising sentiment classification of models trained on the IMDB movie database to classify sentiment about movies [59, 60].

However, as data becomes more accessible and the sentiment analysis techniques become more sophisticated, the performance and scalability of many sentiment analysis tools are constantly improving. This in fact can allow for further advances in the development of *sentiment-aware* dialogue systems, such that dialogue systems can adapt to the dynamics of user sentiment throughout the course of interaction. Depending on the objective function used to optimise, there can be multiple approaches to extract and use the variability in user-sentiment, which can be categorised into two groups:

- **Individual user utterance:** which looks at the sentiment score of individual user utterance, which can offer insight into the specific semantics and vectors of that single interaction such as that found in [58, 59]. This compartmentalised approach allows a deeper evaluation of the content of that one message, whether this is a product, experience or other entity.
- **Contextual user utterance:** examines the thread as a whole can be explored from a temporal perspective, the evolution of the thread, rather than just individual messages [60, 61]. This can give insight as to why the sentiment of the user is going up or down and allows evaluation as to why this is happening. When compared with other threads, trends can be found as to what is causing the fluctuation of sentiment. The difference of sentiment score between the first and last message, which can be referred to as the ‘sentiment swing’ can also be very useful, as this is an example of how the situation has progressed from the perspective of the user.

An example to illustrate user-sentiment swing during dialogue is provided in Tables 3, 4 and 5 which shows three resolved task-oriented interactions. The sentiment score corresponding to the user utterance at each turn is recorded. For simplicity, the variability in user-sentiment at each turn is smoothed in Figure 1. Conversation 1 remains fairly neutral throughout the interaction, ending with a slightly more positive sentiment than at the beginning of the exchange. Conversation 2 shows a positive uptick in sentiment as the relatively simple issue is solved. However, the sentiment of conversation 3 represents the frustration of the user, showing a severe drop in sentiment as they encounter issues with their query. However, as the issue is resolved in the end, the sentiment recovers accordingly to conclude with a positive sentiment score.

Table 3. Conversation One: A thread from the ITS helpdesk dataset.

Source	Utterance	Score	Swing
Conversation One			
User	Hi there, I am unable to copy and paste HTML text - or any text - into Cereus. We have been told by our web editor to paste the text from Word into an online HTML editor and then copy and paste the HTML into Cereus. Unfortunately it doesn't work, even when I right-click to paste, or use control C and V. Thanks,	0.128	
Helpdesk	Are you still having issues with copying and pasting into Cereus via HTML web editor ? What is the name of the Web Editor that gave you this advice ?		
User	Regards Yes I still am having the issues. We use https://html-online.com/editor/ Thank you.	0.27	↑

Helpdesk	I think this might be one of two possible issues. As a first step would you mind using IE11 to access the application via Cereus please? I know sometimes the text editing box can be a bit flaky on newer browsers. Kind regards,		
User	Thank you, but I don't have IE 11. Do you have a safe link you can send me as not sure which source to trust to download. I need IE 11 for Mac...	0.13	↓
Helpdesk	Agh! Sorry —*SR*— I didn't realise you were on a Mac. I don't quite know what to suggest in this case. I haven't heard of anyone else having issues on a Mac but that might be because no one else uses one when trying to use the News app. Cereus is a bit of an old dinosaur and due to be decommissioned soon I'm afraid. I don't suppose you have access to a PC do you? If not I think I will have to put you back to the Help Desk and get them to assign the job to someone who supports Macs. Sorry about this		
User	Hi —*SR*—, I am due to pick up my PC laptop from UWE, but not heard back as to when that could be yet. Plus I need to put out a press release tomorrow morning... Yes, please do put me in touch with one of your Mac guys. Huge thanks for your help though!, Regards	0.24	↑

Table 4. Conversation Two: A thread from the ITS helpdesk dataset.

Conversation Two			
User	Hi BB, Where has the guidance about sign up sheets been moved to?	0	
Helpdesk	Hi, —*SR*— (Request for Information) has been assigned to Learning and Research at the status of 'In Progress'. Open the ticket Thank you		
User	Any news?	-0.12	↓
Helpdesk	Good afternoon *—Person—*, We have had a big clear out of the web site, and are pointing people to the the main support pages for blackboard, if you need further assistance please feel free to contact our help desk. i have found this guidance on sign up sheets here: *—Misc— * thanks		
User	Hey ITS, I might be going blind but where does it mention sign up sheets? I thought multi-sign up sheets were something UWE built? Appreciate your time! *—Person—*	0.36	↑
Helpdesk	Hi *—Person—*, If you are referring to sign up sheets as related to the creation of groups please see the following link: —*Misc*— otherwise if you are referring to the third party 'SignUp Lists' function then the link for that can be found on the above staff guides link page. Regards		
User	That's absolutely grand, thanks —*Person*—!	0.71	↑

Table 5. Conversation One: A thread from the ITS helpdesk dataset.

Conversation Three			
User	Hi Folks, I need to arrange to have 3 laptops (mac or pc) to use for the *—Module—* week at the *—Location*—. How do i go about this please. Groups of students will be planning/editing mixing desk set-ups and making spreadsheets. The desk editing software is a free download Midas M32-Edit software available here *—Misc*— The masterclass runs *—Misc*— to *—Misc*— Cheers, much appreciated!	0.84	
Helpdesk	Hi *—SR*— by *—Person*— for 3 PC or mac laptops for *—Date*— to *—Date*— has been assigned to Client Services Regional - Assignment Details: 3 PC or mac laptops for *—Date*— to *—Date*— Open the ticket Thank you		
User	Hi Folks, I will need to pick up these computers tomorrow for the early start on Monday morning at *—Location*—. Can you tell me where I can collect them from and if the software isn't on them already how we can install it. Many thanks for your help.	0.47	↓
Helpdesk	Hello *—Person*— Sorry but IT Services do not have a stock of loanable laptops. I would suggest trying the FET Project room. If students are using *—Misc*— built laptops off site they will have to log in to them on site beforehand to create their user profile. If software needs installing ask the Project room to liaise with the *—Room*— ITS helpdesk who will assist with this. Regards		
User	Hi *—Person*—, i realise this probably isn't your fault . . . BUT To wait for 7 days to tell me this is a little bit off. Can you understand why I might think that this falls short of reasonable service? I'm not very happy to find out at the last moment something which you might have told me at the start of this week when I would have had time to do something about it!	-0.76	↓
Helpdesk	*—Person*— has turned up trumps with 2 machines .. they will need to be set up with logins that 45 students can use at the *—Misc*— and with this software . . . I will bring them to *—Room*— in a short while for you to action this. The desk editing software is a free download Midas M32-Edit software available here *—Misc*— The masterclass runs 0900 *—Date*— to 1900 *—Date*— Regards		
User	Hi Folks, All sorted now, crisis averted, many thanks! Cheers	0.36	↑

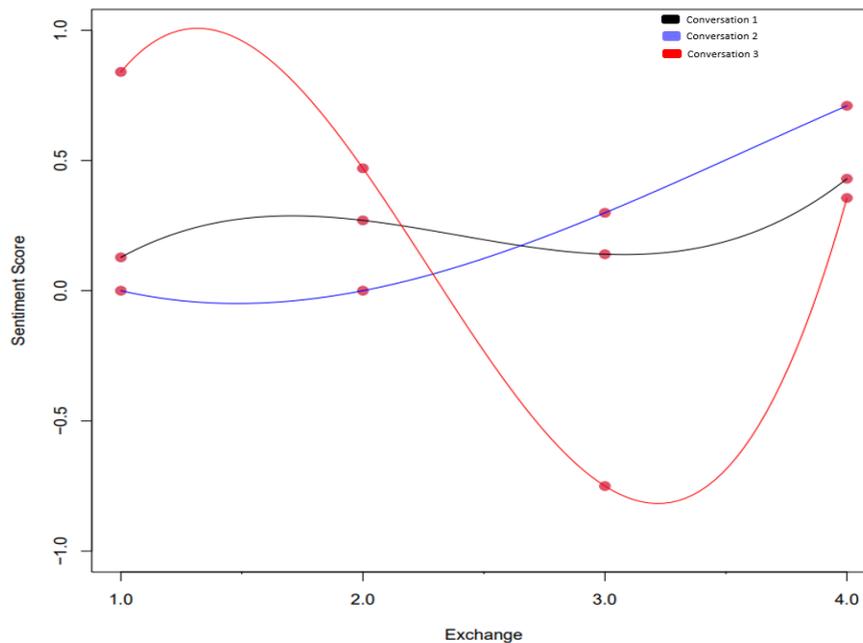


Figure 1. An illustration of the user sentiment score of conversation 1, 2 and 3 from Tables 3, 4 and 5.

Despite the scores fluctuating throughout the interaction, all threads end with a neutral to positive conclusion, indicating that the user was satisfied or happy with the outcome. Whilst this is insightful in itself, the highs and lows provide a chance to understand why these values were exhibited at that point, which could allow for the examination of the objective attributes or semantics used. The values could also just simply be the result of a contextual issue, such as, in this case, a restaurant being fully booked.

However, regardless of the domain in which sentiment analysis is utilised, a cautious apprehension should be taken in interpreting the obtained scores. Modern sentiment analysis tools are advancing, but they are still not mature enough to accurately recognise sarcasm, jokes and nuances of language. There is also the limitation of a lack of distinctive sentiment annotations amongst an already limited amount of datasets readily available, as observed in [60] which subsequently makes it harder to perform accurate analysis on dialogues of a more extensive lexicon.

What's more, sentiment analysis is sensitive to social conventions which are prevalent in human communication. Many interactions through email, for example, will exhibit some form of generic greeting such as 'Good Morning' as well as a sign off (sometimes inserted automatically through a template) such as 'Best Wishes'. These terms are often used by individuals, regardless of the context of their email, which can therefore skew the sentiment score to be higher than the actual substance that their email might elicit.

Therefore, it could be argued that the current state of sentiment analysis makes it a useful tool to gain analytical insight from a corpus of text, but to utilise them as the sole driver for action could potentially lead to erroneous decision making. The context of its usage is important.

2.4. Dialogue Cost

The term ‘dialogue cost’ appears frequently throughout dialogue system literature [62, 63, 64] and typically refers to multiple aspects of resource retrieval and utilisation ranging from the data itself, to the computational power required by the model being utilised. Some literature even refers to the explicit monetary cost of the dialogue system based on the manual labour required to label the data, often using the method of crowdsourcing [65, 66].

Relevant and feature rich data is the foundation for a performing dialogue system, and no matter how good a model is, it cannot compensate for a small or poor quality dataset. Therefore, such resources can be expensive to acquire, whether in terms of time or money [67]. In more domain specific dialogue, the data native to these sometimes unfamiliar domains, plays an even more important role as it highlights semantic and pragmatic phenomena that is unique to that domain.

Alongside task-resolution, dialogue cost is often considered to infer ‘dialogue strategies’ [68] which specify at each stage what the next action to be taken by the system. A dialogue strategy can have the objective of converging towards the goal state in the most efficient way possible through a series of interactions with the user. ‘Efficiency’ can for example, mean access to external resources, the dialogue duration, internal computation time, or resource use. The goal is to reduce these ‘costs’ to allow the system to achieve higher performance.

However, the ambiguity of the term ‘dialogue cost’ can make it a difficult area to assess. The PARADISE framework describes efficiency measures such as the number of turns or elapsed time to complete a task [68, 80, 6], as well as qualitative measures such as inappropriate or repair utterances [70, 71] as potential dialogue costs. Whereas, some researchers explore the term from a reinforcement learning perspective, in which the dialogue cost is a penalisation assigned for taking the wrong action predicated on a pre-defined function. Therefore, it can be a difficult to quantify cost in relation to a dialogue. Even when considering what is typically agreed on, regardless of the context, that dialogue ‘cost’ should be minimised, i.e., to maximise system efficiency, there isn't such established foundation to suggest that, for instance, a shorter —hence more ‘efficient’— dialogue is directly correlated to a better user experience. In fact, it can simply be the opposite.

2.5. Dialogue Cost

The retention rate of a TODS is often referred to as a measure of the number of users that return to use the system within a given time frame. This is another important, yet accessible metric for quantifying dialogue systems' performance. If a company's chatbot aims to replace other communication channels (e.g., lowering call volume), the goal is to obtain significantly higher retention, which can be indicative of higher consumer satisfaction [72]. However, there are plenty of other automated options that allow users to manage accounts easily without speaking to a human. Thus, if a chatbot is focused on customer support, a high retention rate does not necessarily have to be the measure of success [73].

The context and domain in which the TODS is deployed is an important factor to consider when looking at the retention rate of a given dialogue system. If the dialogue system in question is a health-based chatbot for a one-off issue, then the user is unlikely to have to reuse the chatbot, and therefore the metric is less valuable. However, if the chatbot is being deployed as a customer service replacement, then a high retention rate can be interpreted as a positive performance indicator, as it shows the user has enough confidence in the system to reuse it.

Related metrics are those of *dropout* rate and *bounce* rate. The dropout rate refers to the number of users who quit the session with the dialogue system before an outcome had been reached. A high dropout rate for a dialogue system can be a substantial indication of poor performance. The bounce rate is the volume of users that do not utilise the dialogue system for its intended use. A high retention rate with low dropout and bounce rates would suggest a high level of performance.

However, only so much can be derived from the metric of retention rate without some form of user feedback, as the metric is sensitive to anomalies. A dialogue system could perform perfectly, yet a user might not return for other, unknown reasons. This should not be indicative of the performance of the system, yet the metric might suggest this to be the case. Therefore, the larger the set of interactions retention is analysed on, the more insightful the findings will potentially be. Because of this, it could be argued that the rate of retention offers a good overview perspective of system performance, but such considerations should prevent retention rate from being a primary form of performance insight. It is also important to note that the ability to extract the retention rate is not always feasible, as is the case with the ITS helpdesk dataset.

2.6. Response Time Cost

The literature exploring dialogue response time is typically concerned with reducing the time it takes a conversational agent to respond to the user. The consensus is that a user wants responses as quickly as possible, and for the interaction to be as efficient as possible in terms of session time. The focus is often on the mechanics of the model in question, rather than the effect that response time could have on user satisfaction [74].

Alternative studies on response time shift the focus from the desire for instant responses to adding more human-like delays. In their study of using dynamic response delays for machine generated messages, Gnewuch et al [75] prioritise the ‘feel’ of the conversation over speed of response, opting to ‘calculate a timing mechanism based on the complexity of the response and complexity of the previous message as a technique to increase the naturalness of the interaction’. As a result of these dynamic delays, they showed an increase in both the perception of humanness and social presence, as well as a greater satisfaction with the overall dialogue interaction; a faster response time is not necessarily better.

However, as with the majority of the quality attributes, the context and domain are very important to consider. ‘Replika’ [76] is an anthropomorphised chatbot designed as a companion to help battle loneliness. It utilises a slight delay to make the interaction feel more genuine and human-like, as instant replies would make the interaction feel too machine-like and break the social illusion. Conversely, ‘911bot’ [77] is a chatbot that allows a user to describe an emergency situation, and because of this context, any artificial delays would not be appropriate. This highlights the importance of context when considering such conversational attributes to evaluate TODS performance.

Computationally, response time has become much less of a pressing issue in recent times for smaller to medium scale dialogue systems, as abundant computational resources, and innovation in machine learning \ NLP approaches, make instantaneous responses entirely feasible, and as a result, expected. Therefore, it could be argued that whilst a dialogue system might not get praised on its performance for optimal response times, whether instant or timed, it will be negatively graded for sub-optimal response times.

2.6. Conversation Length

The literature exploring the explicit length of conversation is limited. This is due to the fact that the developers predominantly focus on the substance of a message first, with the subsequent message length being as long or short as it needs to be. However, the length of an agent's responses can significantly alter the dynamic of an interaction, as it determines how much information can be conveyed in a single turn. Depending on the topic at hand, if the messages are too short, there is a risk the user will grow frustrated with the lack of detail in the answer, but if the messages are too long, the user's attention may wander.

In their guide to developing "better" chatbots for mental health, Dosovitsky et al [78] argue that "developers should strive to find a module length that enhances intervention fidelity without compromising engagement" and "should focus on creating a few engaging and effective modules at the beginning rather than developing a large variety of untested modules". Simply put, system utterance length should be adaptive, changing relative to the stage of the conversation.

Other work examined the effect of message length relative to the dialogue domain, e.g., [79], emphasising that one of the most important chatbot performance metrics is conversation length and structure. Industry trends suggest aiming for shorter conversations with simple structure, in line with the notion of efficient service. For example, banking chatbots are assumed to provide quick solutions such as sending and receiving money, or checking a balance. When the social aspect of the conversation is more important, fast and concise responses may turn counter-productive.

However, just looking at conversation length from an objective perspective can be misleading. If an analysis is performed in which it is deemed shorter messages are preferred for a given domain, and are subsequently rewarded, then this may undermine the very relevant factor of context. Dialogue systems often have the objective of being as efficient as possible, which would encourage the idea of concise discourse, which may not be a problem. However, some issues and topics simply do not lend themselves to this approach and require further development in the conversation. Therefore, it would be detrimental to the system to simply penalise longer message without any thought to the semantics and context involved. This is not to say conversation length is not a useful quality attribute, as the literature suggests, it is, yet the optimisation of this parameter needs more than just the configuration of a value for utterance length or number of turns.

3. DISCUSSION AND CONCLUSIONS

It is clear that there is no shortage of studies exploring the field of TODS and their performance [80]. However, research into TODS in conjunction with conversational quality attributes, beyond that of task-resolution, are less abundant. One potential reason for this is because these attributes, such as conversation length, response time and user-sentiment are often referred to more as bi-products of the dialogue systems performance in meeting user information requirements.

Although many studies on optimising TODS performance examined metrics for performance evaluation beyond that of task-resolution, thus far, however, the modelling of TODS performance as a multivariate function of multiple conversational quality attributes remains an open question.

Additionally, TODS are still difficult to evaluate. Although there are established methods and frameworks which are frequently referred to in literature, with PARADISE arguably the most applied, yet there is still no standard in place for a novel TODS to be measured against. This is

undoubtedly a hindrance to the field, as it gives a lack of consistency when designing a system and subsequently comparing it with others in the industry. Also, with the growing complexity of modern virtual assistants such as Siri, Bixby and Alexa to name a few, where each could be described as a *sophisticated* TODS, the task of objectively evaluating such systems is only going to become a more complex process.

Therefore, although significant progress has been made in the field of TODS over a relatively short period, there are still various challenges to be overcome. Arguably the most pressing issue is the lack of a standardised protocol for human evaluation, which makes it challenging to compare different approaches to one another [95]. On the other hand, automatic evaluation metrics have proven their utility with their efficiency and undemanding approach to dialogue assessment but are still considered less reliable in comparison to human judgement [132]. A shortage of task-oriented open-source datasets also acts as a bottleneck in the progression of the field, especially when approaching multiple domains. All of which is compounded by a growing expectation of the average user, as TODS are generally becoming more and more innovative on a global scale.

REFERENCES

- [1] P.-H. Su, M. Gasic, N. Mrksic, L. Rojas-Barahona, S. Ultes, D. Vandyke, T.-H. Wen, S. Young, On-line active reward learning for policy optimisation in spoken dialogue systems, arXiv preprint arXiv:1605.07669 (2016).
- [2] O. Vinyals, Q. Le, A neural conversational model, arXiv preprint arXiv:1506.05869 (2015).
- [3] M. Henderson, B. Thomson, J. D. Williams, The second dialog state tracking challenge, in: Proceedings of the 15th annual meeting of the special interest group on discourse and dialogue (SIGDIAL), 2014, pp. 263–272.
- [4] Chatbot market size to reach \$1.25 billion by 2025 — cagr: 24.3%: Grand view research, inc, shorturl.at/gjqwT, accessed: 2021-07-14.
- [5] W. Wang, K. Siau, Trust in health chatbots, Thirty ninth International Conference on Information Systems, San Francisco 2018 (2018).
- [6] M. A. Walker, D. J. Litman, C. A. Kamm, A. Abella, Paradise: A framework for evaluating spoken dialogue agents, arXiv preprint [cmp-lg/9704004](https://arxiv.org/abs/1907.04004) (1997).
- [7] S. Moller, R. Englert, K.-P. Engelbrecht, V. Hafner, A. Jameson, A. Oulasvirta, A. Raake, N. Reithinger, Memo: towards automatic usability evaluation of spoken dialogue services by user error simulations., Ninth International Conference on Spoken Language Processing (01 2006).
- [8] J. D. Williams, K. Asadi, G. Zweig, Hybrid code networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 665–677.
- [9] C. Kamm, User interfaces for voice applications, Proceedings of the National Academy of Sciences 92 (22) (1995) 10031–10037. arXiv:<https://www.pnas.org/content/92/22/10031.full.pdf>.
- [10] M. Walker, J. C. Fromer, S. Narayanan, Learning optimal dialogue strategies: A case study of a spoken dialogue agent for email, in: COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics, 1998.
- [11] N. Fraser, D. Gibbon, R. Moore, R. Winski, Assessment of interactive systems., Mouton de Gruyter, 1998, pp. 564–615.
- [12] J. M. Deriu, A. Rodrigo, A. Otegi, G. Echegoyen, S. Rosset, E. Agirre, M. Cieliebak, Survey on evaluation methods for dialogue systems, Artificial Intelligence Review 54 (1) (2020) 755–810
- [13] J. Schatzmann, B. Thomson, K. Weilhammer, H. Ye, S. Young, Agenda-based user simulation for bootstrapping a pomdp dialogue system, in: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers, 2007, pp. 149–152.
- [14] H. H. Clark, S. E. Brennan, Grounding in communication., Perspectives on socially shared cognition (1991)

- [15] H. H. Clark, *Using language*, Cambridge university press, 1996.
- [16] P.-H. Su, D. Vandyke, M. Gašić, D. Kim, N. Mrkšić, T. H. Wen, S. Young, Learning from real users: Rating dialogue success with neural networks for reinforcement learning in spoken dialogue systems, arXiv preprint arXiv:1508.03386 (09 2015)
- [17] B. Thomson, S. Young, Bayesian update of dialogue state: A pomdp framework for spoken dialogue systems, *Computer Speech Language* 24 (4) (2010) 562–588
- [18] J. Planells, L. Hurtado Oliver, E. Segarra, E. Sanchis, A multi-domain dialog system to integrate heterogeneous spoken dialog systems,
- [19] M. Noseworthy, J. C. K. Cheung, J. Pineau, Predicting success in goal-driven human-human dialogues, in: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, Association for Computational Linguistics, Saarbrücken, Germany, 2017, pp. 253–262.
- [20] C.-S. Wu, A. Madotto, E. Hosseini-Asl, C. Xiong, R. Socher, P. Fung, Transferable multi-domain state generator for task-oriented dialogue systems, in: *ACL*, 2019.
- [21] Y. Huang, J. Feng, M. Hu, X. Wu, X. Du, S. Ma, Meta-reinforced multi-domain state generator for dialogue systems, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, 2020, pp.7109–7118.
- [22] N. Mrksic, D. O Seaghdha, T.-H. Wen, B. Thomson, S. Young, Neural belief tracker: Data-driven dialogue state tracking, in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 1777–1788.
- [23] P. Xu, Q. Hu, An end-to-end approach for handling unknown slot values in dialogue state tracking, 2018, pp. 1448–1457.
- [24] C.-S. Wu, A. Madotto, E. Hosseini-Asl, C. Xiong, R. Socher, P. Fung, Transferable multi-domain state generator for task-oriented dialogue systems, in: *ACL*, 2019.
- [25] J. D. Williams, S. Young, Partially observable markov decision processes for spoken dialog systems, *Computer Speech & Language* 21 (2) (2007) 393–422.
- [26] Z. Zhang, R. Takanobu, Q. Zhu, M. Huang, X. Zhu, Recent advances and challenges in task-oriented dialog systems, *Science China Technological Sciences* (2020) 1–17
- [27] M. A. Walker, D. J. Litman, C. A. Kamm, A. Abella, Paradise: A framework for evaluating spoken dialogue agents, arXiv preprint [arXiv:1907.04004](https://arxiv.org/abs/1907.04004) (1997).
- [28] J. F. Allen, B. W. Miller, E. K. Ringger, T. Sikorski, Robust understanding in a dialogue system, in: *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Vol. 62, 1996, p. 70
- [29] R. Meena, G. Skantze, J. Gustafson, Data-driven models for timing feedback responses in a map task dialogue system, *Computer Speech & Language* 28 (4) (2014) 903–922
- [30] Z. Callejas, R. Lopez-Cozar, Relations between de-facto criteria in the evaluation of a spoken dialogue system, *Speech Communication* 50 (8-9) (2008) 646–665.
- [31] S. M. Robinson, A. Roque, A. Vaswani, D. Traum, C. Hernandez, B. Millspaugh, Evaluation of a spoken dialogue system for virtual reality call for fire training, Tech. rep., University of Southern California Marina Del Rey Ca Inst for Creative Technologies (2007).
- [32] L. Lamel, S. Rosset, J.-L. Gauvain, Considerations in the design and evaluation of spoken language dialog systems (03 2001).
- [33] A. Kamm, M. Walker, D. Litman, Evaluating spoken language systems (06 1999).
- [34] A. Malchanau, V. Petukhova, H. Bunt, Multimodal dialogue system evaluation: a case study applying usability standards, in: *9th International Workshop on Spoken Dialogue System Technology*, Springer, 2019, pp. 145–159.
- [35] E. Raita, A. Oulasvirta, Too good to be bad: Favorable product expectations boost subjective usability ratings, *Interacting with Computers* 23 (4) (2011) 363–371
- [36] V. M., J. Vala, P. Balani, A survey on sentiment analysis algorithms for opinion mining, *International Journal of Computer Applications* 133 (2016) 7–11

- [37] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal* 5 (4) (2014) 1093–1113.
- [38] H. Sinha, A. Kaur, A detailed survey and comparative study of sentiment analysis algorithms, in: 2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS), 2016, pp. 94–98
- [39] R. Prabowo, M. Thelwall, Sentiment analysis: A combined approach, *Journal of Informetrics* 3 (2) (2009) 143–157.
- [40] V. Kharde, S. Sonawane, Sentiment analysis of twitter data: A survey of techniques, *International Journal of Computer Applications* 139 (2016) 5–15
- [41] B. Schuller, G. Rigoll, M. Lang, Hidden markov model-based speech emotion recognition, in: 2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698), Vol. 1, 2003, pp. I–401.
- [42] T. L. Nwe, S. W. Foo, L. C. De Silva, Speech emotion recognition using hidden markov models, *Speech Communication* 41 (4) (2003) 603–623
- [43] L. Devillers, L. Lamel, I. Vasilescu, Emotion detection in task-oriented spoken dialogues, in: 2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698), Vol. 3, 2003, pp. III–549.
- [44] - T. Bui, J. Zwiers, M. Poel, A. Nijholt, Toward affective dialogue modeling using partially observable markov decision processes, in: 1st workshop on Emotion and Computing – Current Research and Future Impact, 2006, pp. 47–50.
- [45] E. Ferreira, F. Lefevre, Expert-based reward shaping and exploration scheme for boosting policy learning of dialogue management, in: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 2013, pp. 108–113.
- [46] S. Ultes, P. Budzianowski, I. Casanueva, N. Mrksic, L. M. Rojas-Barahona, P. hao Su, T.-H. Wen, M. Gai c, S. J. Young, Domain-independent user satisfaction reward estimation for dialogue policy learning, in: INTERSPEECH, 2017.
- [47] J. Shin, P. Xu, A. Madotto, P. Fung, Happybot: Generating empathetic dialogue responses by improving user experience look-ahead (2019). arXiv:1906.08487
- [48] N. Jaques, A. Ghandeharioun, J. H. Shen, C. Ferguson, A. Lapedriza, N. Jones, S. Gu, R. Picard, Way off-policy batch deep reinforcement learning of human preferences in dialog (2020).
- [49] T. Saha, S. Saha, P. Bhattacharyya, Towards sentiment aided dialogue policy learning for multi-intent conversations using hierarchical reinforcement learning, *PLOS ONE* 15 (7) (2020) 1–28.
- [50] J. Acosta, Using emotion to gain rapport in a spoken dialog system., in: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium, 2009, pp. 49–54
- [51] J. Pittermann, A. Pittermann, W. Minker, Emotion recognition and adaptation in spoken dialogue systems, *International Journal of Speech Technology* 13 (2010) 49–60.
- [52] B. Liu, I. Lane, End-to-end learning of task-oriented dialogs, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop, Association for Computational Linguistics, New Orleans, Louisiana, USA, 2018, pp. 67–73
- [53] N. Braunschweiler, A. Papangelis, Comparison of an End-to-end Trainable Dialogue System with a Modular Statistical Dialogue System, in: Proc. Interspeech 2018, 2018, pp. 576–580.
- [54] Q. V. Liao, W. Geyer, M. Muller, Y. Khazaen, Conversational Interfaces for Information Search, Springer International Publishing, Cham, 2020, pp. 267–287.
- [55] E. Ruane, S. Farrell, A. Ventresque, User perception of text-based chatbot personality, in: A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, P. B. Brandtzaeg (Eds.), *Chatbot Research and Design*, Springer International Publishing, Cham, 2021, pp. 32–47
- [56] W. Shi, Z. Yu, Sentiment adaptive end-to-end dialog systems, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 1509–1519.
- [57] T. Saha, S. Saha, P. Bhattacharyya, Towards sentiment aided dialogue policy learning for multi-intent conversations using hierarchical reinforcement learning, *PLOS ONE* 15 (2020)

- [58] H. Kumar, B. Harish, H. Darshan, Sentiment analysis on imdb movie reviews using hybrid feature extraction method., *International Journal of Interactive Multimedia & Artificial Intelligence* 5 (5) (2019).
- [59] A. Yenter, A. Verma, Deep cnn-lstm with combined kernels from multiple branches for imdb review sentiment analysis, in: 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), IEEE, 2017, pp. 540–546
- [60] H. Saif, Y. He, H. Alani, Semantic sentiment analysis of twitter, in: *International semantic web conference*, Springer, 2012, pp. 508–524.
- [61] K. Scheffler, S. Young, Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning, in: *Proceedings of HLT*, Vol. 2, 2002.
- [62] M. A. Walker, D. J. Litman, C. A. Kamm, A. Abella, Paradise: A framework for evaluating spoken dialogue agents, *arXiv preprint cmlg/9704004* (1997).
- [63] J. Relano-Gil, D. Tapias, M. C. Gancedo, M. Charfuelán, L. Hernández, Robust and flexible mixed-initiative dialogue for telephone services, in: *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, 1999, pp. 287–290.
- [64] M. Mitchell, D. Bohus, E. Kamar, Crowdsourcing language generation templates for dialogue systems, in: *Proceedings of the INLG and SIGDIAL 2014 Joint Session*, 2014, pp. 172–180.
- [65] P. Shah, D. Hakkani-Tür, B. Liu, G. Tür, Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning, in: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*, Association for Computational Linguistics, New Orleans - Louisiana, 2018, pp. 41–51
- [66] R. Manuvinakurike, M. Paetzel, D. DeVault, Reducing the cost of dialogue system training and evaluation with online, crowd-sourced dialogue data collection, *Proceedings of SEMDIAL (2015)* 113–121
- [67] E. Levin, R. Pieraccini, W. Eckert, Learning dialogue strategies within the Markov decision process framework, in: 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings, IEEE, 1997, pp. 72–79.
- [68] A. Abella, M. K. Brown, B. Buntschuh, Development principles for dialog-based interfaces, in: *Workshop on Dialogue Processing in Spoken Language Systems*, Springer, 1996, pp. 141–155
- [69] L. Hirschman, C. Pao, The cost of errors in a spoken language system, in: *Third European Conference on Speech Communication and Technology*, 1993
- [70] M. Danieli, E. Gerbino, Metrics for evaluating dialogue strategies in a spoken language system, in: *Proceedings of the 1995 AAAI spring symposium on Empirical Methods in Discourse Interpretation and Generation*, Vol. 16, 1995, pp. 34–39.
- [71] A. Simpson, N. M. Eraser, Black box and glass box evaluation of the sundial system, in: *Third European Conference on Speech Communication and Technology*, 1993.
- [72] M. Dhyani, R. Kumar, An intelligent chatbot using deep learning with bidirectional rnn and attention model, *Materials Today: Proceedings* 34 (2019) 817–824
- [73] A. Nursetyo, E. R. Subhiyakto, et al., Smart chatbot system for e-commerce assistance based on aiml, in: 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, 2018, pp. 641–645.
- [74] M. Kowsher, A. Tahabilder, M. Z. I. Sanjid, N. J. Prottasha, M. M. H. Sarker, Knowledge-base optimization to reduce the response time of bangla chatbot, 2020 Joint 9th International Conference on Informatics, Electronics and Vision and 2020 4th International Conference on Imaging, Vision and Pattern Recognition, ICIEV and icIVPR 2020 (8 2020).
- [75] U. Gnewuch, S. Morana, M. T. Adam, A. Maedche, Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction, in: 26th European Conference on Information Systems: Beyond Digitization-Facets of Socio-Technical Change, ECIS 2018, Portsmouth, UK, June 23–28, 2018. Ed.: U. Frank, 2018, p. 143975.
- [76] Replika. URL <https://replika.ai/>
- [77] J. Martin, 911bot, <https://github.com/surgeforward/911bot> (2016).
- [78] G. Dosovitsky, B. S. Pineda, N. C. Jacobson, C. Chang, M. Escoredo, E. L. Bunge, Artificial intelligence chatbot for depression: Descriptive study of usage, *JMIR Formative Research* 4 (11 2020).
- [79] A. Przegalinska, L. Ciechanowski, A. Stroz, P. Gloor, G. Mazurek, In bot we trust: A new methodology of chatbot performance measures, *Business Horizons* 62 (6) (2019) 785–797.

- [80] Fellows, R., Ihshaish, H., Battle, S., Haines, C., Mayhew, P. and Deza, J.I., 2021. Task-oriented Dialogue Systems: performance vs. quality-optima, a review. *arXiv preprint arXiv:2112.11176*.

© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

EMOJI-BASED FINE-GRAINED ATTENTION NETWORK FOR SENTIMENT ANALYSIS IN THE MICROBLOG COMMENTS

Deng Yang, Liu Kejian, Yang Cheng, Feng Yuanyuan and Li Weihao

Department of Computer Engineering, Xihua University, Chengdu, China

ABSTRACT

Microblogs have become a social platform for people to express their emotions in real-time, and it is a trend to analyze user emotional tendencies from the information on Microblogs. The dynamic features of emojis can affect the sentiment polarity of microblog texts. Since existing models seldom consider the diversity of emoji sentiment polarity, the paper propose a microblog sentiment classification model based on ALBERT-FAET. We obtain text embedding via ALBERT pretraining model and learn the inter-emoji embedding with an attention-based LSTM network. In addition, a fine-grained attention mechanism is proposed to capture the word-level interactions between plain text and emoji. Finally, we concatenate these features and feed them into a CNN classifier to predict the sentiment labels of the microblogs. To verify the effectiveness of the model and the fine-grained attention network, we conduct comparison experiments and ablation experiments. The comparison experiments show that the model outperforms previous methods in three evaluation indicators (accuracy, precision, and recall) and the model can significantly improve sentiment classification. The ablation experiments show that compared with ALBERT-AET, the proposed model ALBERT-FAET is better in the metrics, indicating that the fine-grained attention network can understand the diversified information of emoticons.

KEYWORDS

Sentiment Analysis, Pre-training Model, Emojis, Attention Mechanism.

1. INTRODUCTION

With the rapid development of the Internet, microblog posts have become a platform for young users to express their opinions. Sentiment analysis is the process of analyzing, processing, generalizing and reasoning about subjective texts filled with emotional expression, which has attracted much attention in natural language processing. Microblog comment texts are more informal than ordinary texts, and to analyse the sentiment of microblog comments will generate much practical value. For example, it can be used for e-commerce platforms to conduct microblog marketing and make personal recommendations for users. It can also be used to monitor online public opinion, grasp people's opinions and emotions about social events. In addition, it can understand the public's mental health and identify potential patients with depression nad anxiety.

Traditional methods mainly construct sentiment dictionaries to accomplish the sentiment classification task. Based on manually established seed adjective vocabularies, Hu and Liu [1] proposed a bootstrapping technique by using WordNet to predict the sentiment tendency of opinion words.

Deep learning has achieved good performance in many natural language processing tasks in the past few years. The sentiment classification task has succeeded as a subtask in natural language processing. Zhang [2] conducts recurrent neural networks to obtain word semantic features and word sequence features from sentence vectors and feed them into a softmax classifier to predict the sentiment label of each sentence in Chinese microblogs. Chen [3] et al. combined military sentiment lexicon and BiLSTM model to boost the accuracy and F1-measure. Zhang [4] also used the BiLSTM model to encode semantic information of text, combined with sentiment symbol library to enhance sentiment analysis.

The existence of the same emoji expressing different emotions in different scenarios in Weibo comments complicates the task of sentiment classification. For example, "My stomach hurts, I don't want to talk 🤢" expresses a negative sentiment in the context. A different scenario, "The clothes I ordered arrived and they look beautiful 😊," expresses the exact opposite sentiment compared with the native sentiment polarity in the context. At the same time, the number of emojis impacts the sentiment polarity of the sentence. For example, "Yeah, what you say is totally right 😊😊😊," multiple emojis strengthen the negative emotion expression. Therefore, with the help of sentiment words that co-occur with emojis, we need to extract the necessary textual or contextual features to establish certain connections between emoji and plain text.

Table 1. Microblog comments

emoji	Sentiment	Microblog Comments
🤢	positive	The clothes I ordered arrived and they look beautiful 😊
	negative	My stomach hurts, I don't want to talk 🤢
😊	positive	The weather is good, I feel happy 😊
	negative	Yeah, what you say is totally right 😊😊😊

Most of the existing methods conduct coarse-grained mechanisms to capture interactions between emoji and plain text. If emoji in a complex network environment present emotional polarity diversification or if there are multiple emoji in a sentence. Therefore, this paper proposes a fine-grained attention mechanism to capture the interaction between emoji and plain text. The main problem of the research is to analyze the sentiment tendency information of the microblog text. The main contributions of this paper are summarized as follows:

1. We use ALBERT pre-trained model to learn the word vector of microblog comments. Simultaneously, the model is easy to deploy in engineering due to its fewer parameters.
2. We first adopt emoji2vec to learn bi-sense emoji embeddings and then obtain the inter-emoji embedding as a weighted average of the bi-sense emoji embedding base on the attention mechanism.
3. We propose a fine-grained attention mechanism to extract word-level interaction information between emoji and plain text.
4. Moreover, we design an emoji alignment loss in the objective function to boost the difference of the attention weights towards the emoji which have the same text and different sentiment polarity.

The organization and section of the paper are below:

Section 1: Firstly we introduce the research of the microblog sentiment analysis at home and abroad. Then we briefly analyze the deficiencies of the existing research, and proposed the contribution of the paper.

Section 2: We introduce the research of sentiment analysis with and without emojis in detail.

Section 3: We introduce our model and describe each layer in the model.

Section 4: We adopt our model on the crawled microblog texts and compare it with several state-of-the-art sentiment models.

Section 5 : We summarize our study again and then we offer improvement solutions.

2. RELATED WORK

2.1. Sentiment Analysis

Sentiment analysis is the process of analyzing, processing, generalizing and reasoning about subjective texts filled with emotional expression, which has attracted much attention in natural language processing. Mingjie Ling [5] extracted word representation and sentence position representation from multichannel CNN and LSTM respectively, and experiments showed that the model achieved a better polarity classification ability for Chinese Weibo. Duyu Tang [6] proposed a target-dependent LSTM model where target signals are taken into consideration to boost the classification accuracy.

The earliest application of attention mechanisms is in the field of computer vision. In the literature[7], researchers adopt the attention mechanism on RNN models to implement image classification. Then, Bahdanau et al [8]conducted the attention mechanism to machine translation tasks, which means that the attention mechanism has attracted a lot of attention in the natural language processing. In 2017, the Google machine translation team [9] built the whole model framework with the Attention mechanism to replace the traditional RNN method. It contains a fully connected feed-forward network between each layer in the encoder and decoder structure. Feifan Fan [10] proposed a fine-grained attention mechanism to capture word-level interactions between contexts and aspects in Twitter, and experiments demonstrate that the approach can effectively improve performance.

2.2. Sentiment Analysis with Emoji

Li Nan [11] explored the distribution characteristics and sentiment transformation regularities of emojis in microblog comments from multiple perspectives. The paper classified emojis into high sentiment stability and low sentiment stability based on thresholds and experiment confirmed that emojis can be used in opinion analysis to achieve more accurate sentiment classification. Novak P [12] constructed a sentiment lexicon containing 751 emojis, and experiments demonstrated that comments with emoji expressed more positive sentiments than those without emoji. Pohl [13] investigated the similarity problem of emoji in terms of emoji keywords and emoji embeddings. Experiments verified the model's effectiveness in capturing associations between each emoji. Ben Eisner [14] trained emoji and corresponding descriptive textual information to propose an emoji2vec model for emoji pre-trained embeddings.

Some researchers adopted deep learning to study the sentiment classification task of the emoji-based microblog comments. Zhao [15] adopted CNN and RNN networks based on attention mechanism to extract semantic features and weighted the sentiment tendency values of the text and emojis to predict the sentiment tendency of Chinese microblog comments. Felbo [16] predicted the appearance of emoji with pre-training deep neural networks, which is effective to extract emotion information from emoji in sentiment classification and sarcasm detection task. Li [17] conducted a convolutional neural network to predict the occurrence of emoji and learn emoji embedding jointly through a matching layer based on cosine similarity. These approach, in our

context, adopt emoji as independent inputs to predict the sentiment label, which suffer from ignoring the interactivity between emoji and plain text.

Lou [18] combined attention mechanisms to measure the contribution of each word in sentiment polarity based on emojis, although the approach cannot effectively handle microblog comments containing multiple types of emojis. Yuan X[19] proposed an emoji-based collaborative attention network to learn the interactive sentiment semantics of text and emoji. The model feed the text vector, text-based emoji vector, and emoji-based text vector into the convolutional neural network to predict the sentiment polarity of microblog texts. Experimental results show that the method outperforms several baselines for microblog sentiment classification. Chen [20] combines a more robust and fine-grained bi-sense emoji embedding to represent complex semantic and sentiment information effectively. An attention-based mechanism of the LSTM network selectively attend on the relevant emoji embeddings to understand rich semantics and sentiment better. The experiments on the Twitter dataset demonstrate the model outperforms the state-of-the-art models.

3. THE PROPOSED MODEL

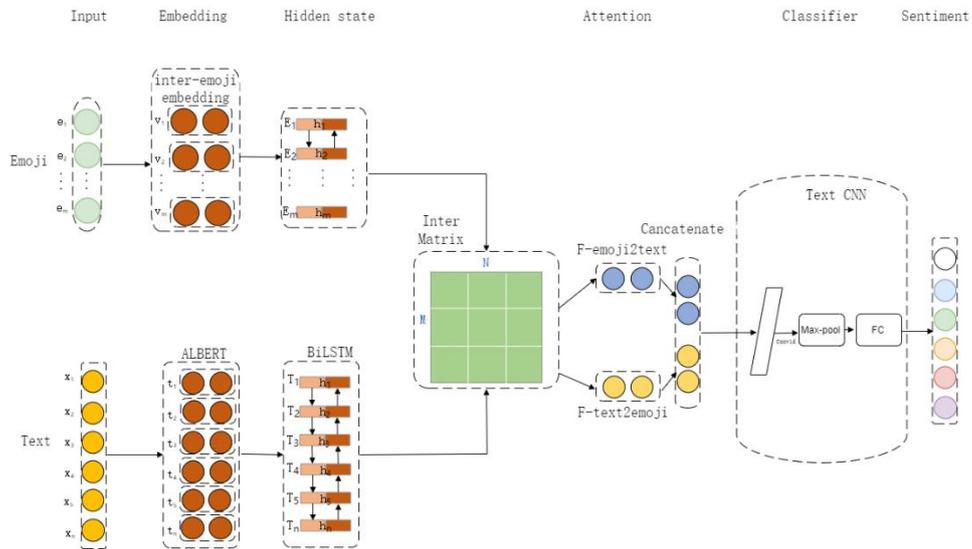


Figure 1. ALBERT-FAET Model

The main problem to be solved in this study is the analysis of the emotional information of barrage comments. For an given sentence $L=\{x_1,x_2,x_3,\dots,x_n,e_1,e_2,\dots,e_m\}$, where n represents the number of text words and m represents the number of emoji words. The sentiment polarity corresponds to 'positive' and 'negative' sentiments of a reviewer. We present the overall architecture of the proposed Emoji-based Fine-grained Attention Network model in Figure 4. It consists of the embedding layer, hidden layer, attention layer, and textCNN classifier layer. Firstly, we use a attention-based network to learn inter-emoji embedding and we adopt ALBERT to learn word embedding. Then we feed the embeddings into the BiLSTM network to capture the temporal interactions among words and propose a fine-grained attention mechanism to describe word-level interactions between emojis and text. Finally, the concatenated vector are fed into a CNN classifier to predict the sentiment label of the microblog comments.

3.1. Embedding Layer

3.1.1. Text

The text adopts ALBERT to learn word embedding. ALBERT utilizes factorized embedding parameterization, cross-layer parameter sharing, and sentence order prediction(SOP) strategies to deepen the model while reducing parameters, to achieve better results than the BERT models in various natural language processing tasks.

3.1.2. Emoji

We first assigned two distinct tokens to each emoji, one is the specific emoji used in a positive sentimental context, and the other is the emoji used in a negative sentimental context. Each token is embedded into a different vector using emoji2vec to obtain a bi-sense embedding \cite{chen2018twitter} to each emoji. We first learn text embedding using ALBERT and obtain inter-emoji embedding by a simple attention network between plain text and emojis.

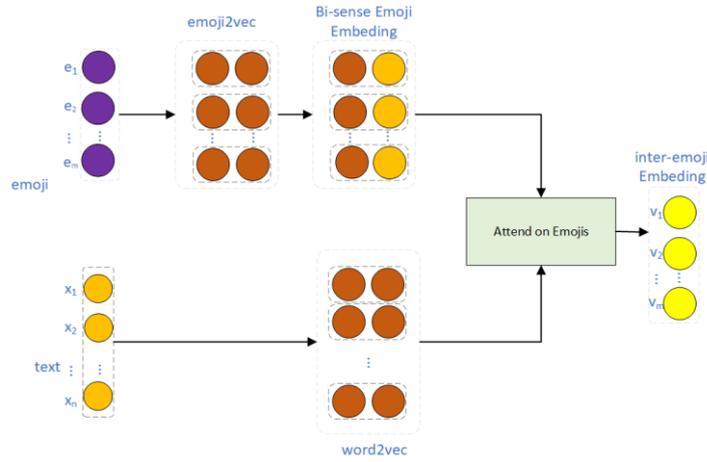


Figure 2. Emoji Processing

$$u_{m,i} = f_{att}(e_{m,i}, w_m) \quad (1)$$

$$\alpha_{t,i} = \frac{\exp(u_{t,i})}{\sum_{i=1}^m \exp(u_{t,i})} \quad (2)$$

$$v_t = \sum_{i=1}^m (\alpha_{t,i} \cdot e_{t,i}) \quad (3)$$

$i \in (1, u)$ in $e_{m,i}$ denotes the i -th sense embedding of the emoji, $f_{att}(\cdot, w_{att})$ denotes the attention function that is based on the current word embedding, represents the attention weight, and α_t represents the attention weight, and $e_{m,i}$ represents the inter-emoji embedding.

We feed the text and emoji into the embedding layer, the output sequence is $L' = \{t_1, t_2, \dots, t_n, v_1, v_2, \dots, v_m\}$

3.2. Hidden Layer

We adopts bidirectional Long Short-Term Memory Network (BiLSTM) to capture the temporal interactions among words. The operations in an LSTM unit for time step t is formulated in Equation below:

$$\mathbf{i}_t = \sigma(W_i \mathbf{x}_t + U_i \mathbf{h}_{t-1} + b_i) \quad (4)$$

$$\mathbf{f}_t = \sigma(W_f \mathbf{x}_t + U_f \mathbf{h}_{t-1} + b_f) \quad (5)$$

$$\mathbf{o}_t = \sigma(W_o \mathbf{x}_t + U_o \mathbf{h}_{t-1} + b_o) \quad (6)$$

$$\mathbf{g}_t = \tanh(W_c \mathbf{x}_t + U_c \mathbf{h}_{t-1} + b_c) \quad (7)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \quad (8)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (9)$$

where \mathbf{h}_t and \mathbf{h}_{t-1} represent the current and previous hidden states, \mathbf{x}_t denotes the current LSTM input, \mathbf{W} and \mathbf{U} denote the weight matrices. Then, we extract deep semantic meaning from the sequence got from the embedding layer to obtain the semantic-rich feature vector $\hat{L}^c = \{\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_n, \mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_m\}$, where T_i is the feature vector of the first i text word after processing, and E_i is the feature vector of the corresponding emoji.

3.3. Attention Layer

In this paper, we propose a fine-grained attention mechanism to describe word-level interactions between emojis and text and to evaluate how emojis affect the overall sentence sentiment polarity.

Formally, we define an interaction matrix $\mathbf{U} \in \mathbb{R}^{N \times M}$ to describe the interaction between an emoji E and a text T , where U_{ij} denotes the interactivity between the i th text word and the j th emoji. The interaction matrix \mathbf{U} is computed by the following equation :

$$U_{ij} = W_u([\mathbf{E}; \mathbf{T}_j; \mathbf{E}_i * \mathbf{T}_j]) \quad (10)$$

$W_u \in \mathbb{R}^{1 \times 6d}$ denotes the weight matrix, $[\cdot]$ denotes the vector concatenation across row, $*$ denotes element wise multiplication, and then we use \mathbf{U} to compute the attention vectors in both directions.

3.3.1. F-Emoji2Text estimates which emoji should pay more attention to and are hence critical for determining the sentiment. We can compute the attention weights e^{ie} by

$$s_i^{fe} = \max(U_{i,:}) \quad (11)$$

$$s_i^{fe} = \max(U_{i,:}) \quad (11)$$

$$e_i^{ie} = \frac{\exp(s_i^{fe})}{\sum_{k=1}^N \exp(s_k^{fe})} \quad (12)$$

where s_i^e obtains the maximum similarity across column. And then we can get the attended vector as follows:

$$\mathbf{m}^{fe} = \sum_{i=1}^N e_i^{fe} \cdot E_i \quad (13)$$

3.3.2. F-Text2Emoji estimates which text should pay more attention to and are also critical for determining the sentiment. We can compute the attention weights

\mathbf{t}^{ft} by

$$\mathbf{s}_i^{ft} = \max(U_{i,:}) \quad (14)$$

$$\mathbf{t}_i^{ft} = \frac{\exp(\mathbf{s}_i^{ft})}{\sum_{k=1}^N \exp(\mathbf{s}_k^{ft})} \quad (15)$$

Then we use an average pooling layer on \mathbf{t}^{ft} to get the attended vector $\mathbf{m}^{ft} \in \mathbb{R}^{2d}$:

$$\mathbf{m}^{ft} = \text{Pooling}([\mathbf{t}_1^{ft}, \dots, \mathbf{t}_i^{ft}]) \quad (16)$$

Finally, we concatenate fine-grained attention vectors as the final representation $\mathbf{m} \in \mathbb{R}^{4d}$

$$\mathbf{m} = [\mathbf{m}^{ft}; \mathbf{m}^{fe}] \quad (17)$$

3.4. Text CNN Layer

We take the concatenated vector into a CNN classifier to predict the sentiment label of the microblog comments.

We adopt $[w_1, w_2, \dots, w_c]$ to denote the set of filter kernels in the convolution operation and then map the input $V \in \mathbb{R}^{d \times c}$ to a new feature map $U \in \mathbb{R}^{d' \times c'}$.

3.5. Model Training

The existing methods train each text word and emoji separately, and seldom consider the fine-grained interaction between emoji and text. Experiments show that the fine-grained interaction between emoji and text can bring additional valuable information. Therefore, a text alignment loss function is proposed in the paper. The text is constrained by the alignment loss, and each text word will focus on the more important emoji by comparing with other text words.

$$d_{io} = \sigma(\mathbf{W}_d[\mathbf{T}_i; \mathbf{T}_o]) \quad (18)$$

$$\ell_{\text{align}} = - \sum_{i=1}^{M-1} \sum_{o=i+1}^M \sum_{k=1}^N d_{ij} \cdot (x_{ik}^{fe} - x_{ok}^{fe})^2 \quad (19)$$

In particular, for text words \mathbf{x}_i and text words \mathbf{x}_o . The paper calculate the square loss on the fine-grained attention vectors \mathbf{x}_i^{fe} and \mathbf{x}_o^{fe} , and also estimate the distance d_{io} between \mathbf{x}_i and \mathbf{x}_o as the loss weight. Where σ is the sigmoid function, \mathbf{W}_d is the weight matrix for computing the distance, x_{ik}^{fe} and x_{ok}^{fe} are the attention weights on k-th context word towards text word \mathbf{x}_i and \mathbf{x}_o , respectively.

4. EXPERIMENTS

In the paper, we crawled 60,000 microblog comments of length greater than 5 using API as the original experimental data. After removing the text excluding emoji from the microblogs and cleaning the data, the corpus are labelled for sentiment polarity by using a manual approach. Finally, we get a dataset consisting of 8930 texts containing emojis. Among them, 4418 were positive texts and 4512 were negative texts. In the paper, the texts are divided into training, validation and test based on 7:2:1. In detail, the training set includes 6250 sentences, the test set includes 1786 sentences and the validation set includes 894 sentences. The distribution of text containing emojis is shown in the following table:

Table 2. The statistics of the microblog datasets

corpus	positive	negative	total
Training	3092	3158	6250
Testing	884	902	1786
Validation	442	452	894
Total	4418	4512	8930

4.1. Evaluation Indexes

In the experiment, the evaluation indexes that have been commonly used in NLP tasks were adopted, and they were as follows: Precision (P), Recall (R), Accuracy (Acc), and F1 values, and they were respectively calculated by:

$$precision = \frac{TP}{TP + FP} \quad (20)$$

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (22)$$

Precision (P) indicates the proportion of samples correctly predicted. In other words, precision measures quality. Recall (R) represents the proportion of samples wrongly predicted. The f1-score is a number between 0 and 1, contributing to the measurement of precision and recall by calculating the harmonic mean of them. Accuracy (Acc) represents the ratio between correctly predicted samples and the total number of samples, and it is a more global index.

4.2. Hyperparameters

In our Experiments, the hidden state d is set to 200, and the dropout ratio is set to 0.2 during the training period. The batch size is 64, the number of iterations is 10, and the maximum length of the text is limited to 100. The Adam optimizer was used to optimize model parameters, and the learning rate is initialized to 5×10^{-4} . We randomly split the microblogs into the training, validation and test sets in the proportion of 7:2:1. The whole framework was built and trained by PyTorch.

4.3. MODEL Performance on Microblog Texts

For this experiment, we test several state-of-the-art sentiment models on our dataset of microblog text:

emoji2vec [14] contains 1661 embeddings, trained on Unicode descriptions of emojis, to improve natural language processing tasks that previously used word2vec to learn word embeddings.

LSTM [17] (long short-term memory) is widely adopted in natural language processing tasks. It controls the transmission state using gate units and selectively memorizes information to process sequence tasks.

TextCNN [22]+word learns word embeddings through word2vec and feeds them into CNN networks to extract semantic features.

TextRCNN [23]+word replaces the convolutional layer with a bidirectional recurrent layer, and the concatenated vector is fed into the classifier to complete the classification compared with TextCNN.

ET-BiLSTM [24] is an emojis-enhanced sentiment analysis model. The model contains the contextual information of the sentence into emoji to learn emoji-based auxiliary representation of the comments.

BERT [21] is a pre-trained model proposed by Google in 2017, which adopts a masked language model to the bidirectional transformer to finish pre-training tasks. Then the last few layers of model parameters need to be fine-tuned to achieve satisfactory results.

BERT+emoji2vec uses BERT to learn the word vector of text and employs emoji2vec to learn emoji embedding. The model concatenates emoji embeddings and text embeddings to complete the sentiment classification.

Table 3. The performance comparisons of different models on the microblogs

Model	Acc	Micro-P	Micro-R
Emoji2vec	0.658	0.642	0.660
TextCNN+word	0.742	0.741	0.743
ET-BiLSTM	0.821	0.819	0.823
BERT	0.802	0.814	0.806
BERT+emoji2vec	0.832	0.837	0.831
ALBERT-FAET	0.852	0.855	0.856

Compared with emoji2vec, TextCNN+word and BiLSTM+word models achieve good results in the sentiment classification task due to the neural networks' robust feature extraction ability. The results of the BERT model are generally better than those pre-trained with word2vec, indicating that BERT can capture deeper text features. Comparing BERT and BERT+emoji2vec, experiments show that emoji information can improve indexes effectively, indicating that emoji information can improve model performance. In addition, the ALBERT-FAET model proposed in this paper outperforms the previous benchmark on Chinese microblog comments. On the one hand, The model assigns two distinct tokens to the emoji to obtain the bi-sense emoji embedding, and a text-based self-attention mechanism is adopted to learn inter-emoji embeddings. On the

other hand, the model proposes a fine-grained attention mechanism to capture the word-level interaction between emoji and text, which brings additional effective information.

Table 4. The performance comparisons of ALBERT-FAET variants

Model	Acc	Micro-P	Micro-R
ALBERT-AET	0.842	0.840	0.845
ALBERT-FAET	0.852	0.855	0.856

Table 5. Prediction results of ALBERT-FAET and ALBERT-FAET partial examples

Number	Microblog Text	ALBERT-FAET	ALBERT-AET	True Label
(1)	My stomach hurts and I don't want to talk 😞	negative	negative	negative
(2)	It's a good day and I feel happy 😊	positive	positive	positive
(3)	My favorite weather 😊 🌸 🌺	positive	positive	positive
(4)	My ordered clothes arrived and they look beautiful 😊	positive	negative	positive
(5)	The favorite East King 😊, hung himself	negative	positive	negative
(6)	Yeah,What you say is relatively right 😊 😊 😊	negative	positive	negative

ALBERT-AET adopt a simple coarse-grained interaction between emoji and text to learn emoji features, and then concatenates text features to finish the sentiment classification. ALBERT-FAET defines an interaction matrix to describe the word-level interaction between emoji and text. Emoticons in (1), (2), and (3) have the same sentiment polarity as text in the current context, so ALBERT-AET, a coarse-grained attention mechanism, can correctly predict the sentiment polarity of the microblog text. The emoticons in (4) and (5) show inconsistency between emotion polarity and text polarity, and multiple emoticons appear in (6). At this time, the ALBERT-AET prediction results are very different from the real results. And ALBERT-FAET can accurately identify the sentiment polarity of the microblog text, which indicates that the fine-grained attention network can learn the dynamic feature information in the emoji, which is helpful to improve the sentiment classification accuracy of the model.

5. CONCLUSION

In this paper, we propose an emoji-based fine-grained attention network for microblog sentiment analysis. Specially, we propose an cross matrix to analyze the word-level interactions between text and emojis, which may bring extra valuable information. More importantly, we design a text alignment loss in the objective function to enhance the difference of the attention weights towards the text which have the same emoji and different sentiment polarities. The model achieves top results for sentiment classification on the crawled microblog dataset.

More fine-grained information can also be considered for sentiment analysis of the microblog comments. For instance, the sender's name of the microblog text, the time when the microblog text was sent mat taken into consideration. Besides, we may alter the finer granularity of the microblog text. We can label the microblog comments into five catorgories: very positive, positive, neutral, negative, and very negative.

In the future, we consider combining the multi-dimensional features of microblogs for sentiment analysis, such as considering the personality characteristics of microblog users.

ACKNOWLEDGEMENTS

We thank Ban Minchao, Zhao Zhiguo and the anonymous reviewers for their comments.

REFERENCES

- [1] Hu M, Liu B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004: 168-177.
- [2] ZHANG Yangsen,JIANG Yuru,TONG Yixuan.Study of Sentiment Classification for Chinese Microblog Based on Recurrent Neural Network[J].Chinese Journal of Electronics,2016,25(04):601-607.
- [3] Chen L C, Lee C M, Chen M Y. Exploration of Social Media for Sentiment Analysis Using Deep Learning[J].Soft Computing, 2020, 24(11): 8187-8197.
- [4] Zhang Y, Zheng J, Huang G, et al. Microblog Sentiment Analysis Method Based on A Double Attention Model[J]. Journal of Tsinghua University Science and Technology, 2018, 58(2): 122-130.
- [5] M. Ling, Q. Chen, Q. Sun and Y. Jia, "Hybrid Neural Network for Sina Weibo Sentiment Analysis," in IEEE Transactions on Computational Social Systems, vol. 7, no. 4, pp. 983-990, Aug. 2020.
- [6] Tang D, Qin B, Feng X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 3298-3307.
- [7] Mnih, Volodymyr, Heess, Nicolas, Graves, Alex, et al. Recurrent Models of Visual Attention[J].,2014.
- [8] Bahdanau, Dzmitry et al. "Neural Machine Translation by Jointly Learning to Align and Translate[J]." CoRR abs/1409.0473 (2015): n. pag.
- [9] Vaswani, Ashish et al. "Attention is All you Need[J]." ArXiv abs/1706.03762 (2017): n. pag.
- [10] Fan F, Feng Y, Zhao D. Multi-grained attention network for aspect-level sentiment classification[C]//Proceedings of the 2018 conference on empirical methods in natural language processing. 2018: 3433-3442.
- [11] Li Nan and Zhang Yuhui.An Analysis of Web Opinion Combining the Dynamic Characteristics of Emoji[J].Modern Intelligence,2021,41(08):98-108.
- [12] Kralj Novak P, Smailović J, Sluban B, et al. Sentiment of emojis[J]. PloS one, 2015, 10(12): e0144296.
- [13] Pohl H, Domin C, Rohs M. Beyond just text: semantic emoji similarity modeling to support expressive communication[J]. ACM Transactions on Computer-Human Interaction (TOCHI), 2017, 24(1): 1-42.
- [14] Eisner B, Rocktäschel T, Augenstein I, et al. emoji2vec: Learning Emoji Representations from their Description[C]//Conference on Empirical Methods in Natural Language Processing. 2016: 48.
- [15] Zhao Xiaofang and Jin Zhigang.Multi-dimensional sentiment classification of microblog based on Emoticons and short texts[J].Journal of Harbin Institute of Technology,2020,52(05):113-120.
- [16] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann.2017. Using millions of emoji occurrences to learn any-domain representations fordetecting sentiment, emotion and sarcasm. EMNLP 2017, September 9-11, 2017. 1615–1625.
- [17] Li X, Yan R, Zhang M. Joint emoji classification and embedding learning[C]//Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data. Springer, Cham, 2017: 48-63.
- [18] Yinxia Lou, Yue Zhang, Fei Li, Tao Qian, and Donghong Ji. 2020. Emoji-Based Sentiment Analysis Using Attention Networks. <i>ACM Trans. Asian Low-Resour. Lang. Inf. Process.</i> 19, 5, Article 64 (August 2020), 13 pages.
- [19] Yuan X, Hu J, Zhang X, et al. Emoji-Based Co-Attention Network for Microblog Sentiment Analysis[C]//International Conference on Neural Information Processing. Springer, Cham, 2021: 3-11.

- [20] Chen Y, Yuan J, You Q, et al. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM[C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 117-125.
- [21] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [22] Chen Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.
- [23] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]//Twenty-ninth AAAI conference on artificial intelligence. 2015.
- [24] Zhang J, Li X, Du Y, et al. A Deep Learning Model Enhanced with Emojis for Sina-Microblog Sentiment Analysis[C]//2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS). IEEE, 2019: 236-242.

AUTHORS

Deng Yang was born in Xichang, Sichuan, China, in 1993. He is currently pursuing the bachelor's degree in computer science and software engineering, Xihua University. His research interests are mainly natural language processing technology, including sentence level sentiment analysis and aspect level sentiment analysis.

A BIG DATA DRIVEN SYSTEM TO IMPROVE RESIDENTIAL IRRIGATION EFFICIENCY USING MACHINE LEARNING AND AI

Kai Segimoto¹, Nelly Segimoto¹ and Yu Sun²

¹Arcadia High School, 180 Campus Dr, Arcadia, CA 91006

²California State Polytechnic University, Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

California has been prone to drought; starting in 2011, there were 376 consecutive weeks of drought [1]. More effective tools are necessary to combat water scarcity, in particular in irrigation systems [3]. This paper designs an application to modify current water-saving techniques to create a more environmentally friendly irrigation system [2]. We developed a Big Data Driven System to Improve Residential Irrigation Efficiency. Our design uses the raspberry Pi controller based on an IoT system with a database connected to the cloud. We designed a mobile app to interact with the system and collect the data and a machine learning algorithm to analyze and generate recommendations based on the given data [4]. We applied our application to the irrigation systems of California Residents and conducted a qualitative evaluation of the approach. The results show that trend-based water saving techniques were effective in reducing water usage without sacrificing the health of the plants being irrigated.

KEYWORDS

Data mining, Cloud computing, Machine Learning, IoT system.

1. INTRODUCTION

For centuries, California weather records have documented intense droughts. With the further commercial and agricultural development, water scarcity has become ever prevalent and increasingly damaging to the environment as well as well-being of residents [5]. As life-long California residents, we have seen the environmental consequences of these droughts and the ineffectiveness of our water-saving techniques. Since all signs indicate that drought will persist, finding ways to reduce water usage is imperative. One of the most inefficient usages of water can be seen in residential irrigation systems. Because of the large consumption of water in irrigation systems, in every city in California, utility companies have urged residents to limit water usage by issuing mandates limiting the dates and times residents are permitted to water the plants and fining those who fail to adhere to these rules. Unfortunately, this method isn't effective -- especially since most water usage goes to irrigating lawns, and gardens in residential homes need to be watered based on weather conditions. It is also unfortunate that California residents have been slow to shift to more drought-tolerant gardens. This has led to the importation of water from nearby states -- a costly and environmentally damaging practice. In recent years, technological developments have led to the creation of irrigation systems that can be controlled virtually [6]. With the creation and adoption of irrigation systems that irrigate based on the needs of plants, California residents will be able to minimize water usage by utilizing it as efficiently as possible, thus conserving water and benefiting not only the environment, but also their finances.

In California, there are many pre-established water-saving techniques. Many tend to vary on the location of the city and the people in charge of water distribution, but in general, most residential neighborhoods set in place guidelines dictating when it is permissible to water plants [7]. These times tend to be two or three times a week, during hours of the day generally ranging from 8pm to 8am, or after dark and before sunrise. Water companies charge residents fees for not abiding to these guidelines, but this method of water regulation is hardly beneficial to reducing the amount of water usage. While initially, these guidelines may seem effective, the issue with this method is that it doesn't take into account weather conditions nor how long plants are being watered. First, it is ineffective because people already water their plants every other day, so regulating which days those are does not reduce water usage. Second, a set frequency during all seasons is not effective due to heat and precipitation being large factors in the amount of water a plant needs, and third, the time frame is so wide that many still end up over irrigating their plants due to lack of knowledge on how long their plants should be watered for. Modernized sprinkler systems are often left unadjusted to match weather conditions because the way they are constructed is not user friendly, leading to many older adults (usually the ones who can afford large, residential homes) being unable to adjust their sprinkler systems to current weather conditions. This is detrimental because sprinkler settings may be set matching summer weather, leading to an extreme use of water during winter periods that need virtually no water. Other smart sprinkler systems run into many of the same issues because individuals are unaware of how long they need to water their plants and lawns, and although they can easily control their sprinklers virtually, the lack of this knowledge leads to over-watered lawns and plants and thus a waste of valuable water [8].

In this paper, our goal is to create an irrigation system driven by data mining and machine learning algorithms to effectively control the sprinkles to save the water [9]. Our method is inspired by the Classic IoT control system and data drive recommendation systems.

We began by building the device that measured weather conditions. We connected the temperature and humidity sensor to a single board computer, Raspberry pi. The Raspberry pi sends the weather data to the Firebase database, which connects to the app we are coding called Smart Irrigation. The app then displays the real-time temperature and humidity. Using Python on the Jupiter Notebook, we utilized data that recorded the conditions in 36 cities in CA over five years to train a model with a machine learning algorithm that predicts the temperature and humidity of the upcoming week. When the app receives the temperature and humidity from the Firebase, the algorithm suggests whether or not the sprinkler system should be turned off, and from the app, the sprinkler system can be shut off.

There are some good features of our smart irrigation system [10]. First, we collected data from 2013 - 2020 in California to train the model, which is suitable to predict future California weather conditions and generate recommendations. Second, the Raspberry Pi provides stable control to the power of sprinkles which can reduce the cost and improve the production. Third, the mobile app is easy to access and operate, expanding the production influence. Therefore, we believe that the smart system we built can effectively solve the problem and help save water in California.

To demonstrate how the irrigation system works we used simulation software to show when the sprinklers start compared to the weather and the result of a lawn. The goal of the simulation was to show how using the sprinkler system you save water and money. By comparing the efficiency and cost effectiveness of the irrigation system to using normal sprinklers or manually watering. The efficiency can be measured by comparing the water wasted and time spent to the result of how the lawn looks. After comparing the irrigation systems, the results are almost the same but using our irrigation system was far superior. The time spent and the water wasted manually watering a lawn was way higher than the automatic systems by a large margin. Both our system

and other smart irrigation systems had much greater efficiency with lower water wastage, but ours was slightly better. Other smart irrigation systems check the humidity of the soil and past weather patterns, but run on a timer so it isn't saving as much water as possible. Our system checks the humidity and past weather patterns, but only turns on during the morning when watering is most effective and on days when watering is necessary without the lawn degrading. Because of this, it is slightly more effective at saving as much water as possible. Also, our system is far more cost efficient since instead of buying a hundred dollar control unit, one can just buy a few pieces of hardware and download an app.

The rest of the paper is organized as follows: Section 2 details the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we outlined in Section 2; Section 4 presents the relevant details about the experiment we conducted, followed by the related works discussed in Section 5. Finally, Section 6 gives the concluding remarks, as well as discussing the future of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. How to use Python and Raspberry Pi

One challenge that we faced was learning how to learn to use python and how to use the Raspberry Pi. Since I was unfamiliar with Python Programming, I had to learn how python worked and what the imported modules did. Figuring out which things to import and how to use them was difficult since I had never imported modules or used their functions. Another problem that we encountered was how to use the Raspberry Pi. Since we were new to using Raspberry Pi's everything from connecting it to the WiFi to connecting it to the Firebase was new. Everything we tried always somehow had small problems that we ran into that were very tedious to fix. For example, the IP address for some reason kept changing every time we restarted the Raspberry Pi, so we had to figure out how to keep it from changing. Connecting the circuit to the Raspberry Pi was also new, but I figured it out in the end.

2.2. How to connect the flutter app with the server

At first, we created a flutter app to run the program but it would not connect with the server. Because of this we had to change it to a thunkable app and a Firebase database. The thunkable interface was different and we faced many challenges when using the code, since many of the functions did not appear. The instructions put on the website did not appear to help the problem, but after a while we eventually figured out how to use the functions connecting the code to the firebase. Creating the program that could take values from the Firebase and display it on an app was also a challenge since we didn't know what most of the functions did. After researching what the things did, we were able to create a code to update the app from the readings in the Firebase and display it on the screen in real time.

2.3. How to find data and make it usable

Another challenge we faced was in finding data and making our data usable for the machine learning algorithm. First, most of the resources we found only included temperature data, but since we were looking for both temperature and humidity together, all of those resources were unusable. When we finally found data that included temperature and humidity, we found that the numbers produced results almost opposite of what we expected-- during the day the temperature

was cooler than during the night. This was definitely wrong, so we investigated and found that all the data was recorded with the same time zone, regardless of what time it actually was in the area being observed. After figuring out what time zone the data was recorded in, we needed to make all of the data fit the actual time in the area, the final challenge in making the data usable for the machine learning algorithm was in writing code so that all of the cities we were using would have the correct time correlation to the weather and humidity.

3. SOLUTION

SmartIrrigation is a smart Irrigation control system based on an IoT system and driven by big data [14]. The settings of the irrigation system can be controlled through the mobile app. The customers can get suggestions from the cloud based on the big data computing result to drive the on/off setting of the system.

SmartIrrigation integrates with the database, mobile app, and machine learning cloud computing server.

The main control system uses a raspberry pi embedded device with a temperature sensor and a humidity sensor to detect weather conditions. The data will be sent to the Firebase database and will sync up to the mobile app that was developed with flutter with a machine learning algorithm. To train the algorithm, we analyzed data from 36 cities from California for the past 7 years. We analyzed everything from the most general overview down to the details to provide accurate suggestions that apply to California cities.

Therefore, it can be used for detecting the weather conditions, predicting the temperature and humidity of certain days based on data, and performing analysis tasks such as whether to control the sprinklers by person or by algorithm to save water.

Additional relevant tasks are to manage all the systems besides one city. The main technical challenge of the system is managing the uploading and storing data —while delivering the real-time tem and hum data necessary for cloud computing. Furthermore, the accurate rate of the machine learning algorithm must be more than 95% to make the prediction stable. To achieve these goals, our tool consists of three main components (see the thick boxes in Figure 1):

- a hierarchical data structure for storing the tem and hum attributes in an aggregated format [15];
- an stable cloud server with trained data to improve the efficient
- a powerful mechanism for efficiently turning on and off the controller.

We also provide a set of navigation techniques for exploring the graph. The following sections describe these components in detail.

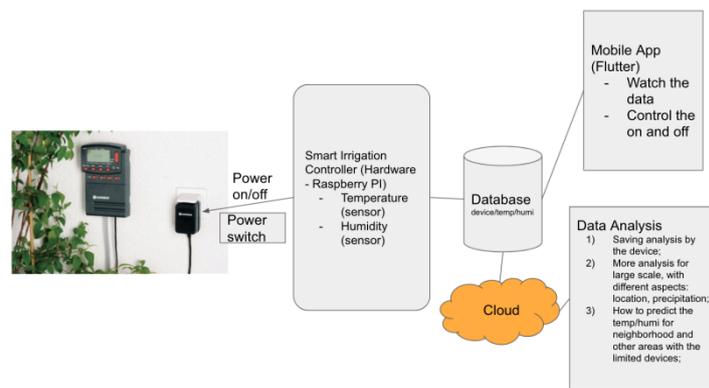


Figure 1. Overview of the system

We created a Flutter app to monitor and receive the data from the sensor, here are 3 screens we have:

1. The first screen is the login screen, which is used to login to find the data from your sensor.



Figure 2. Screenshot of login page

```
margin: EdgeInsets.only(top:10, left:20, right: 20),
child: RaisedButton(
  shape: RoundedRectangleBorder(
    borderRadius: BorderRadius.circular(30),
    side: BorderSide(color: Colors.blue)), // RoundedRectangleBorder
  textColor: Colors.white,
  child: Text("Login"),
  onPressed: () {
    print("Email:");
    print(emailController.text);
    print("Password:");
    print(passwordController.text);
    FirebaseAuth.instance.signInWithEmailAndPassword(
      email: emailController.text.trim(), password: passwordController.text).then(
        print("Login Successful");
        print(val.toString());
        print("userid is");
        print(val.user.uid);
        Navigator.push(context, MaterialPageRoute(builder: (context) => MyHomePage
      }).catchError((error) {
        print("error");
        print(error.toString());
      });
  });
```

Figure 3. Screenshot of code (1)

2. The second one is the Homepage Screen, which loads the data..... and will update the data....
The 3 buttons function is listed below:

- 1) Refresh Button: Refresh the data from pi and sensor
- 2) ON/OFF: Control the Pi by app
- 3) Get Suggestion: Get Suggestion from Server to turn on or off



Figure 4. Screenshot of Home page


```

Container(
  width: 400,
  height: 40,
  margin: EdgeInsets.only(top:10, left:20, right: 20),
  child: DropdownButton<String>({
    value: temRange,
    icon: Icon(Icons.arrow_drop_down),
    iconSize: 40,
    elevation: 16,
    style: TextStyle(color: Colors.green),
    underline: Container(
      height: 20,
      color: Colors.green
    ), // Container
  }, onChanged: (String newValue) {
    setState(
      () {
        temRange = newValue;
      });
  },
  items: <String>["Higher than 95", "85-95", "70-85", "50-70", "Lower than 50"]
    .map<DropdownMenuItem<String>>((String value) {
      return DropdownMenuItem<String>(value: value, child: Column(
        children: [
          SizedBox(height:5),
          Text(value)
        ],
      )); // Column // DropdownMenuItem
    //Sizedbox(height:5), Text(value));
    })
    .toList()
  ), // DropdownButton
), // Container

```

Figure 7. Screenshot of code 3

4. EXPERIMENT

To give a better suggestion, we use the python algorithm and put it on a suggestion response server. We analyze the data from 2013 - 2016 and give the suggestion based on what we learned on the curve.

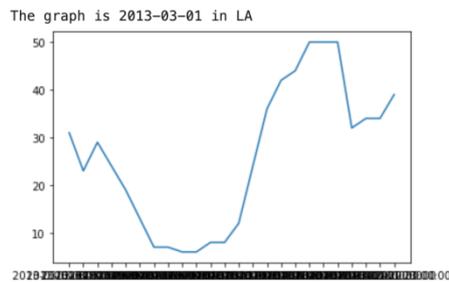


Figure 8. A Hum changes in one day of LA

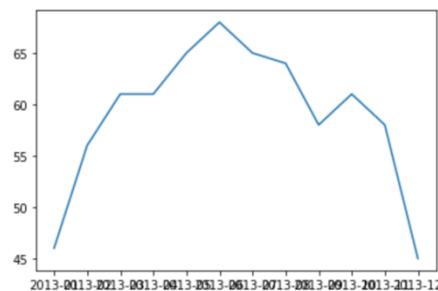


Figure 9. A Average Hum changes in one Year of LA

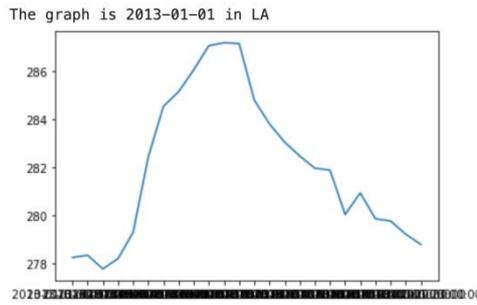


Figure 10. Temperature changes in one day of LA

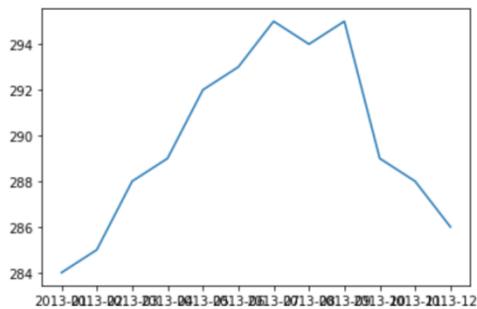


Figure 11. An Average Temperature changes in one Year of LA

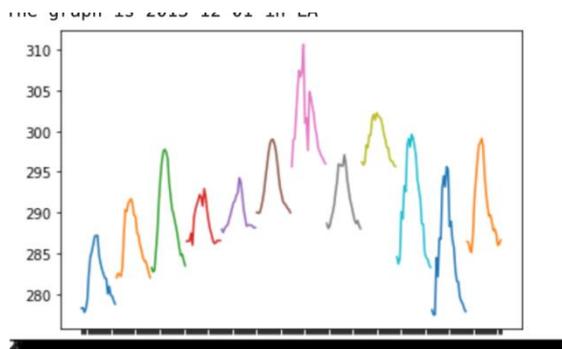


Figure 12. An Average Potential temperature changes by months

Based on these graphs we can determine when it is best to use water to conserve as much water as possible and save more money. The data we do the analysis comes from the website of the US Statistics. Based on the above analysis, we have selected a total of 5 different methods for irrigation recommendations. To suggest by A Hum changes in one day of LA, Average Hum changes in one Year of LA, Average Temp changes in one Year of LA, Average Hum changes in one Year of LA, Potential tem changes by months. Result shows that suggestion based on both Average Temp changes in one Year of LA and Average Hum changes in a day of LA have the highest prediction accuracy. So our algorithm used both Average Temp changes in one Year and Average Hum changes in a day as the data to train the suggestion model.

5. RELATED WORK

On the Smart Irrigation System, written by S. Darshna, T.Sangavi, Sheena Mohan, A.Soundharya, Sukanya Desikan, they create a system that monitors the amount of soil moisture and temperature using a predefined range of soil moisture and temperature that can be varied with soil or crop type [10]. The watering system can be turned on or off in case the moisture or temperature of the soil deviates from the desired range. When the soil is dry or has a high temperature, it will activate the irrigation system to pump water and bring the soil back into the desired range.

Kriti Taneja; Sanmeet Bhatia created an automated irrigation system using Arduino sensors to efficiently utilize water during irrigation [11]. The system has a soil moisture sensor for the soil near the plants and a water level sensor in a water container where water will be pumped to the plants. They created an algorithm using maximum values of soil moisture to control the quantity of water in the soil and the water level sensor to measure the amount of water being sent from the tank. Using an Arduino with an ATmega328 micro controller, they can use automatic irrigation to turn the pumping motor in the tank on or off depending on the dampness of the soil. By doing so, it eliminates human intervention and saves more water by efficiently and effectively irrigating the plants. The micro controller collects values from the soil sensors and depending on the values, it pumps water out. A LCD screen is connected to the micro controller to display the values from the soil and water pump. The water level sensor is used so that the water tank always contains enough water to efficiently irrigate the crops.

In a paper written by KK Namala, Krishna Kanth Prabhu A V, Anushree Math, Ashwini Kumari, Supraja Kulkarni, they propose a smart irrigation system that can be used to control the watering of plants, so their humans can be less human intervention [12]. They focus on the wastage of water and saving as much as possible, which is a problem in modern times. It also helps save time, is cost effective, protects the environment, and is low maintenance with a low operating cost which results in an efficient irrigation service. The Raspberry Pi is used to make the system compact and sustainable. It uses a sensor to measure the moisture of soil and based on the desired moisture it can switch a relay that controls a solenoid valve for irrigation.

6. CONCLUSIONS

We have created a temperature and humidity app that detects factors such as temperature and humidity to decide if it is a good or bad time to use water [12]. This app can be used to conserve as much water as possible during the California drought. By saving water it not only helps out the world's water problem and it saves as much money on water bills as possible.

One variable that affects the accuracy is that the temperature data is limited to past values in the Los Angeles area. This means that the data is only accurate in Los Angeles and other areas for the most part, inaccurate. This makes it unusable in other areas across the country and the globe. Also, the current device is pretty impractical. It must be connected to a power source and have a WiFi connection. This isn't practical as the device has to be outside to measure temperature and humidity values. Also, using raspberry pi is quite expensive and along with all the other components the cost adds up.

To solve the limitations many things can be done. The historical data can be expanded to many major cities across America or the World, and in time it will be able to accurately function in all cities [13]. Another solution to the practicality is by using an Arduino instead of a Raspberry Pi. Since the device only performs one function, an Arduino would be better suited as it is only able to run one program. This could also bring the cost down.

REFERENCES

- [1] Mishra, Ashok K., and Vijay P. Singh. "A review of drought concepts." *Journal of hydrology* 391.1-2 (2010): 202-216.
- [2] Blanke, Amelia, et al. "Water saving technology and saving water in China." *Agricultural water management* 87.2 (2007): 139-150.
- [3] Darshna, S., et al. "Smart irrigation system." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* 10.3 (2015): 32-36.
- [4] Joorabchi, Mona Erfani, Ali Mesbah, and Philippe Kruchten. "Real challenges in mobile app development." 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. IEEE, 2013.
- [5] Barrett, Christopher B., Michael R. Carter, and C. Peter Timmer. "A century-long perspective on agricultural development." *American Journal of Agricultural Economics* 92.2 (2010): 447-468.
- [6] Van Lente, Harro. "Promising technology: The dynamics of expectations in technological developments." (1995): 0741-0741.
- [7] Shamir, Uri Y., and Charles DD Howard. "Water distribution systems analysis." *Journal of the Hydraulics Division* 94.1 (1968): 219-234.
- [8] Thompson, Allen L., et al. "Testing of a water loss distribution model for moving sprinkler systems." *Transactions of the ASAE* 40.1 (1997): 81-88.
- [9] Romero, Cristobal, and Sebastian Ventura. "Data mining in education." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3.1 (2013): 12-27.
- [10] Darshna, S., et al. "Smart irrigation system." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* 10.3 (2015): 32-36.
- [11] Taneja, Kriti, and Sanmeet Bhatia. "Automatic irrigation system using Arduino UNO." 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2017.
- [12] Yamazoe, Noboru, and Yasuhiro Shimizu. "Humidity sensors: principles and applications." *Sensors and Actuators* 10.3-4 (1986): 379-398.
- [13] Simonton, Dean Keith. "Qualitative and quantitative analyses of historical data." *Annual review of psychology* 54.1 (2003): 617-640.
- [14] Buhl, Hans Ulrich, et al. "Big data." *Business & Information Systems Engineering* 5.2 (2013): 65-69.
- [15] Tanimoto, Steven, and Theo Pavlidis. "A hierarchical data structure for picture processing." *Computer graphics and image processing* 4.2 (1975): 104-119.

FUNREADING: A GAME-BASED READING ANIMATION GENERATION FRAMEWORK TO ENGAGE KIDS READING USING AI AND COMPUTER GRAPHICS TECHNIQUES (FOR SPECIAL NEEDS)

Jiayi Zhang¹, Jiayu Zhang¹, Justin Wang² and Yu Sun³

¹Northwood High School, 4515 Portola Pkwy, Irvine, CA 92620

²The Peddie School, 201 S Main St. Hightstown, NJ 08520

³California State Polytechnic University, Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

Children with ASD or ADHD are having a hard time learning and understanding, and there's no perfect education system [1][2]. However, audios and animations can improve their reading effectiveness. This paper designs an application to have animated characters talking with audio based on the text using Optical Character Recognition, text to speech, and Natural Language Processing.

KEYWORDS

AI, Computer Graphics Techniques, Machine Learning.

1. INTRODUCTION

Children with ASD/ADHD are having a hard time learning and understanding the text. In the educational industry, there are usually two traditional teaching methods specifically targeting kids with ADHD or ASD [3]. The first method is establishing specific schools for students with ADHD or ASD [4]. This method also has a flurry of disadvantages. First, it is costly. Establishing a school requires millions of dollars. Second, it cannot help kids with ASD or ADHD learn how to communicate with other people since those kids are confined inside these special schools and cannot interact with others [5]. In conclusion, the general existing offerings or treatments for ADHD or ASD are ineffective, inefficient, and financially burdensome. The second method is video games. These attract kids' attention and were thought to effectively combat ADHD. Teachers integrate knowledge inside games letting students learn through video games. However, this teaching method has drawbacks. First, children usually only focus on the game itself rather than the knowledge behind the game. Second, video games have a proven addictive nature in their development. Thus, video games as a treatment are low in efficiency and not long-term feasible in full-time education. This paper develops an application to automatically convert traditional storybooks to interactive videos to help educate and treat children with ADHD or ASD [6].

EndeavorRx is the first and only clinically proven prescription video game that participates as a treatment for ADHD children [7]. The game is designed to improve attention control by strengthening focus, interference processing, and multitasking through an immersive video game experience.

Studies show that video stories have vast potential for helping with the academic, social, and emotional education of children with ADHD and autism. Therefore, we developed a product that automatically converts traditional storybooks to interactive videos to help educate and treat children with ADHD or ASD.

Our treatment is more efficient and feasible than games or specialized schools. First, our solution is scalable. Video game programmers need lots of time to design the plot games [8]. Thus, the production of video games is generally slower, so it takes more time for customers to get benefits from the product. On the contrary, our treatment is converting stories into videos at the click of a button. Most crucially, the conversion of stories does not need time, because the plot was already designed by the story's author. Second, our solution is more specific and targeted because our product is designed to custom for each student. Every user has a unique background and story recommendation based on their interest. Third, our solution is more effective. We teach kids with ADHD or ASD knowledge in areas of math, science, and language arts while incorporating features engaging their communication functions through interaction functions. Kids can interact with AI or community members. Lastly, our proposal is a more accessible education method than traditional schools since we present less opportunity cost to parents, students, and government organizations. Our products are more scalable, effective, and affordable than the traditional methods of educating students with neurodivergence.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met while creating this application and the further development we think about and trying to achieve; Section 3 focuses on the details of our current solutions; Section 5 shows related work; Section 6 gives the conclusion remarks and the future development of the application.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Selecting Model (Apply SALSA)

Choosing a model is extremely important to the first impression and the willingness to use the application for users. These steps become challenging as not every model could cooperate with the SALSA Lipsync. Besides the animation and lip sync, the outfit of the character is also important for children to develop interest.

2.2. Using API flexibly

While building the app, it is important to have API apply and connect to the systems. In order to minimize the resources used in the app, we have to use the API flexibly and reduce space and resources.

2.3. Developing more interactions

We want to create a precise mental health system for the app and would like to cooperate with professional psychologists and psychiatric clinics to develop an algorithm that evaluates the

user's condition to accomplish real-time tracking and optimization of content presentation, as well as the animation, and push media according to the user's preferences. For now, we are developing more interactions to allow users to receive more reactions. Making the app flexible to ask the right questions is also a risk, and the online video platform along with the group chat system we will build in the future is challenging.

3. SOLUTION

Story Book is an application that provides users with an efficient, productive, and cost-effective product to automatically create animated videos from pictures of written material. The image is captured using the camera on our users' mobile devices. We implemented the google cloud vision service into the C# code on unity using API to complete our OCR phase [9]. To turn the text output from OCR into speech, we need to process a human voice simulation. We used IBM's online Watson text-to-speech service the same way we used the google cloud service to complete the text-to-speech function. We employed spaCy's pre-trained NLP model to extract keywords from the text such as places and organizations and organize them in a text file. Up to this point, all the information we need is ready, so the application starts to create the video. We need background images and a character for the videos. We used Ultimate Stylized Business Women to create a 3D figure. This 3D model allows us to control the muscles of the character and thus control their lip movements and facial expressions, which are both essential for the character to look natural during the video [10]. We also utilized the SALSA LipSync Suite asset to automatically match the lip movements of the character with the audio file. We have a library of pictures of different topics that can act as backgrounds. Using the keywords that we detected with NLP, we find pictures in that category from our library and set these pictures as interchangeable backgrounds for the video. Finally, after all these processes, a full video with the same context and meaning as in the written material is created.

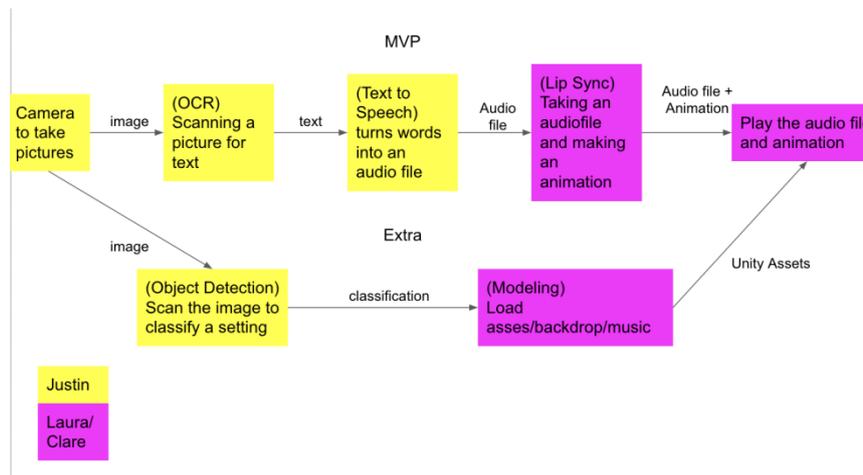


Figure 1. Overview of the solution

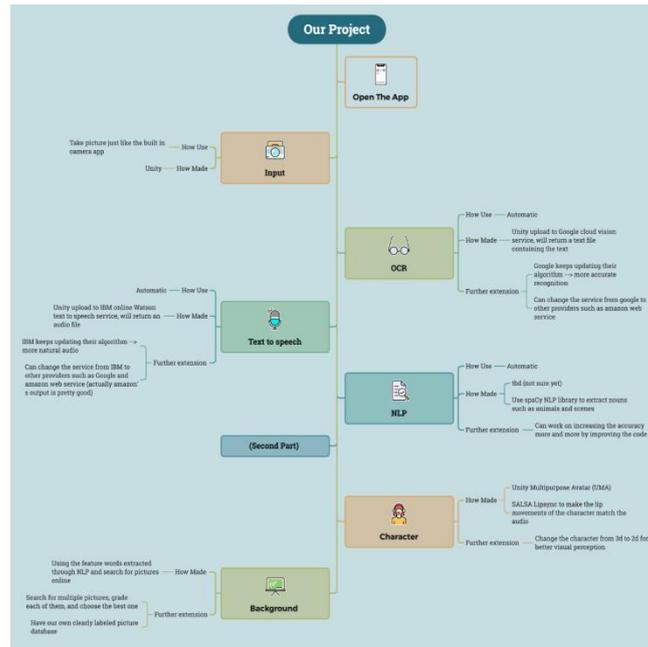


Figure 2. Project structure

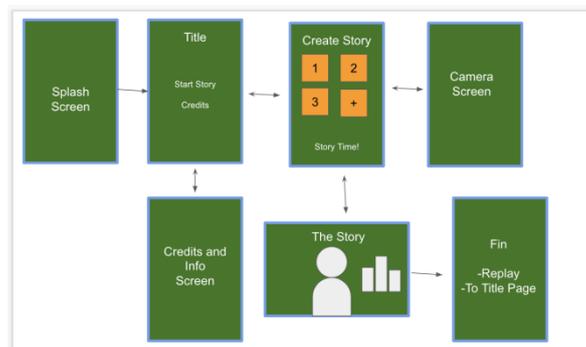


Figure 3. Overview of the APP



Figure 4. Screenshot of the APP

The application is developed using Unity 3D. Unity 3D is a cross-platform 3D engine for creating 3D games and applications for mobile, desktop, the web, and consoles. As shown in the picture

above, the app has 3 main screens, title, create a story, and animation. These screens are different scenes in Unity and are connected using the LoadScene method. In the Create Story scene, the “+” sign allowed the users to take photos of the text materials. The camera is connected to the OCR, after the user takes the photo, it would automatically recognize the text and transform an audio file.

```
public IEnumerator ProcessImage()
{
    _status = OCR_Status.OCR_STARTED;
    string imagePath = Path.Combine(Application.persistentDataPath, ImageName);

    WWWForm form = new WWWForm();

    byte[] bytes = File.ReadAllBytes(imagePath);
    form.AddBinaryData("image", bytes, ImageName, "image/png");

    UnityWebRequest www = UnityWebRequest.Post(google_fucntion_url, form);
    www.timeout = Timeout;

    yield return www.SendWebRequest();

    if (www.result != UnityWebRequest.Result.Success)
    {
        Debug.Log(www.error);
        _status = OCR_Status.OCR_ERROR;
    }
    else
    {
        Debug.Log(www.downloadHandler.text);
        _status = OCR_Status.OCR_SUCCESS;
        _text = www.downloadHandler.text.Replace("\n", " ").Replace("\r", " ");
    }
}
}
```

Figure 5. Screenshot of code 1

```
// Watson Code section
private string Authenticate(string username, string password)
{
    string auth = username + ":" + password;
    auth = System.Convert.ToBase64String(System.Text.Encoding.GetEncoding("ISO-8859-1").GetBytes(auth));
    auth = "Basic " + auth;
    return auth;
}

public IEnumerator ProcessAudio()
{
    _status = OCR_Status.SPEECH_START;

    Debug.Log("Started");
    string method = @"POST/synthesize";
    string urlmethodpath = watson_fucntion_url + method;

    string jsonText = "{\"text\":\"" + Text + "\"}";
    byte[] jsonBytes = System.Text.Encoding.UTF8.GetBytes(jsonText);

    Debug.Log(
    Debug.Log(

    string aut

    WWWForm fc
    //form.Adc

    UnityWebRequest www = UnityWebRequest.Post(urlmethodpath, form);

    www.SetRequestHeader("Authorization", authorization);
    www.SetRequestHeader("Content-Type", "application/json");
    www.SetRequestHeader("Accept", "audio/mp3");

    www.uploadHandler = (UploadHandler)new UploadHandlerRaw(jsonBytes);

    string path = Path.Combine(Application.persistentDataPath, $"{ImageName}.mp3");
    www.downloadHandler = new DownloadHandlerFile(path);

    Debug.Log("Sent Request!");

    yield return www.SendWebRequest();

    if (www.result != UnityWebRequest.Result.Success)
    {
        Debug.Log("Some error occurred in the Post Request");
        Debug.Log(www.result);
        Debug.Log(www.error);
        Debug.Log(www.responseCode);

        foreach (KeyValuePair<string, string> pair in www.GetResponseHeaders())
        {
            //print(pair.Key + " " + pair.Value);
        }

        _status = OCR_Status.SPEECH_ERROR;
    }
    else
    {
        Debug.Log("Success maybe in the Post Request");
        _status = OCR_Status.SPEECH_SUCCESS;
    }
}
}
```

Figure 6. Screenshot of code 2

As shown in the image above, we utilized the google cloud vision service to extract text from pictures with OCR and the IBM Watson text-to-speech service to generate audio of simulated human voice based on the text outputted from OCR. These two functions are closely connected so that the text will get delivered to the Watson online service immediately to generate audio files.

We based our NLP algorithm on the spaCy NLP library, which provides advanced natural language processing in Python [11]. With spaCy, we can filter out specific nouns, verbs, or both nouns and verbs in a specific orientation.

The animation and background are built with assets downloaded from the Unity assets store, the items contain their own codes that are able to animate them in Unity. The background is built in different scenes using the assets downloaded, while the application processes the image, it would automatically select the background with the keywords extracted from the text using the NLP algorithm. The character is animated with the SALSA LipSync asset and the audio file generated from the Watson online service, allows the characters to have lip movements according to the text material. After the application finishes processing and clicking the “star” button, it would load the StoryScene, and the selected background and character would be in that scene.

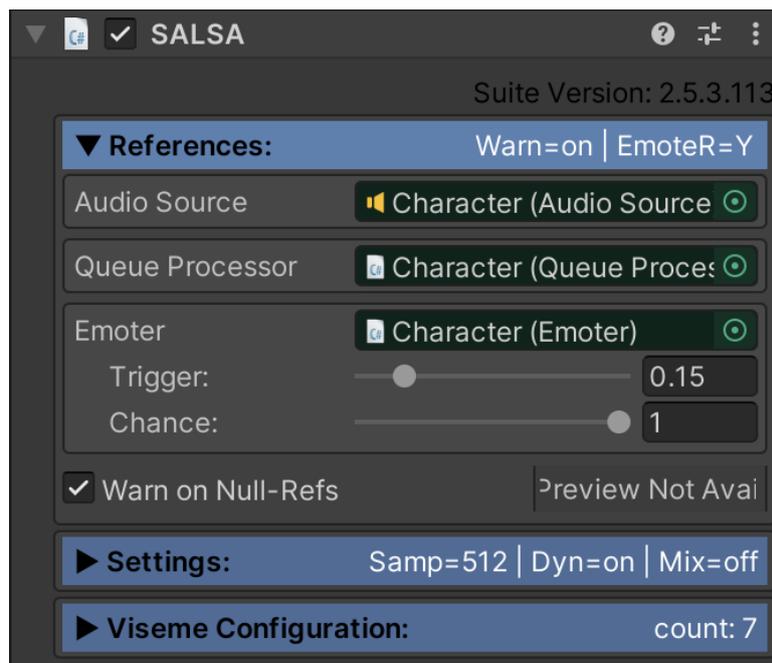


Figure 7. Screenshot of SALSA

4. EXPERIMENT

4.1. Experiment 1

In order to verify that our solution can effectively solve problems at different levels and have good user feedback, we decided to select multiple experimental groups and comparison groups for several experiments. For the first experiment, we want to prove that our solution works stable and continuously, so we choose a group size of 100 different NLP inputs text sentence different kind of target. The goal of the first experiment is to verify if IBM chat bot could analyze all the text sentence if the AI read the sentence right works good for all the target sentence. Experiments have shown that almost all targets in different types tested the right result. Moving enemy has the most correct rates, which means our user are works more better in aiming the moving enemy in the game. This experiment could explain that the data sets do have a obvious impact on the finding the targets and aim it, because the data we are using have a high rates of the moving enemy. The average correct rate of 100 different types of the aims shows below:

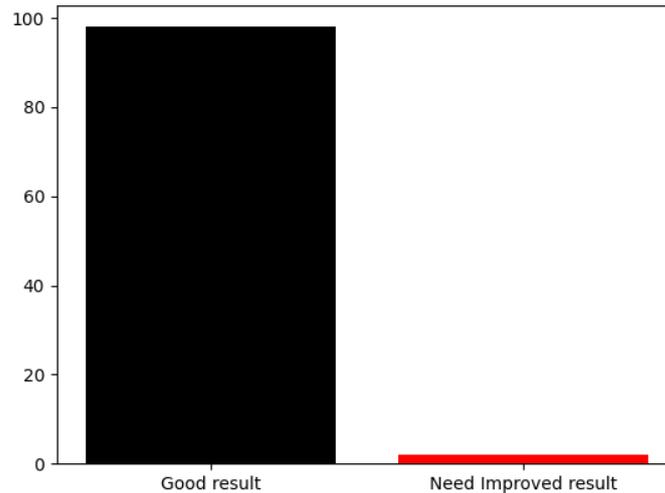


Figure 8. Number of good and need improved result

5. RELATED WORK

Hayashi, Masaki, et al developed technology that generates CG animation generated automatically from a text-based script, it uses TVML, is a language with script structure used in actual TV production, to edit the text and FIL to complete text to speech [12]. Their project is similar to ours with the automatically generated animation from text. The technology of Hayashi, Masaki uses different language templates.

Large, Andrew, et al conclude that animation helps children to identify the major plots more successfully than reading the text through his experiment [13]. Their project also uses animation to create solutions for social problems, but is designed for special science experiments.

Shaw, Rebecca, and Vicky Lewis reported ADHD children to focus more and give more accurate responses to the animated computerized tasks [14]. Different from our project, this work is an investigation and study on the positive impact on education of computers to ADHD children, and our project is to implement the help.

6. CONCLUSIONS

The project provides users with an efficient, productive, and cost-effective product to automatically create animated videos from pictures of written material; we design it to help educate and treat children with ADHD or ASD. The current prototype can automatically create videos from images of written sources. The user merely needs to click on the “start” button on our homepage and start taking pictures of books, our application handles everything behind the stage. After a few seconds, the users would be able to watch the animations. The most important characteristic of this product is its ability to incorporate multiple functions and design them to work together at the fastest speed possible. All parts of our project are already implemented into different commercialized applications and services. A large portion of our application’s functions uses pre-existing services in our databases. These include Google’s online vision service to achieve optical character recognition (OCR), IBM’s Watson text-to-speech service to generate audio files, Unity’s Multipurpose Avatar with SALSA LipSync Suite asset to match the

characters' lip movements with audio data, and the spaCy library to accomplish natural language processing (NLP) [15].

We still have a lot more proposed functions in development, including the interactive feature in videos, an online video sharing platform, and a group chats feature. The interactive feature in videos is aimed to help children with ADHD/ASD concentrate and learn communication skills. The online video sharing platform is akin to YouTube with the difference that our users are not video creators. All videos on this sharing platform are automatically generated through our application. When users use our application to generate a video for themselves, they are able to choose whether to upload this video to our sharing platform so that others without the original written material can also access it. The group chat feature is implemented so that users can join specific learning group chats and share related videos and support each other. This feature also helps people with ADHD/ASD set up their own virtual community.

REFERENCES

- [1] Jensen, Peter S., David Martin, and Dennis P. Cantwell. "Comorbidity in ADHD: Implications for research, practice, and DSM-V." *Journal of the American Academy of Child & Adolescent Psychiatry* 36.8 (1997): 1065-1079.
- [2] Kasari, Connie, et al. "Social networks and friendships at school: Comparing children with and without ASD." *Journal of autism and developmental disorders* 41.5 (2011): 533-544.
- [3] Sonuga-Barke, Edmund JS, et al. "Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments." *American Journal of Psychiatry* 170.3 (2013): 275-289.
- [4] Reichow, Brian, et al. "Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD)." *Cochrane database of systematic reviews* 10 (2012).
- [5] McAlister, Alfred L., Cheryl L. Perry, and Guy S. Parcel. "How individuals, environments, and health behaviors interact." *Health Behavior* 169 (2008).
- [6] Moody, Amelia K., Laura M. Justice, and Sonia Q. Cabell. "Electronic versus traditional storybooks: Relative influence on preschool children's engagement and communication." *Journal of Early Childhood Literacy* 10.3 (2010): 294-313.
- [7] Dondlinger, Mary Jo. "Educational video game design: A review of the literature." *Journal of applied educational technology* 4.1 (2007): 21-31.
- [8] Rahimi, Farzan Baradaran, et al. "A Game Design Plot: Exploring the Educational Potential of History-Based Video Games." *IEEE Transactions on Games* 12.3 (2019): 312-322.
- [9] Nguyen, Trong Duc, et al. "Exploring API embedding for API usages and applications." 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE). IEEE, 2017.
- [10] Jain, Vishal. "3D model of attitude." *International Journal of Advanced Research in Management and Social Sciences* 3.3 (2014): 1-12.
- [11] Vasiliev, Yuli. *Natural Language Processing with Python and SpaCy: A Practical Introduction*. No Starch Press, 2020.
- [12] Hayashi, Masaki, et al. "T2V: New Technology of Converting Text to CG Animation." *ITE Transactions on Media Technology and Applications* 2.1 (2014): 74-81.
- [13] Large, Andrew, et al. "Multimedia and comprehension: The relationship among text, animation, and captions." *Journal of the American society for information science* 46.5 (1995): 340-347.
- [14] Shaw, Rebecca, and Vicky Lewis. "The impact of computer-mediated and traditional academic task presentation on the performance and behavior of children with ADHD." *Journal of Research in Special Educational Needs* 5.2 (2005): 47-54.
- [15] Joorabchi, Mona Erfani, Ali Mesbah, and Philippe Kruchten. "Real challenges in mobile app development." 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. IEEE, 2013.

A REAL-TIME MULTIPLAYER FPS GAME USING 3D MODELING AND AI MACHINE LEARNING

John Zhang¹ and Yu Sun²

¹Crean Lutheran High school, 12500 Sand Canyon Ave, Irvine, CA 92618

²California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

AIs have been a key component in the gaming industry throughout its history. Developers have had multiple ways of creating new AI models that best suit their game in order to enhance the playing experience. However, with the increase in the popularity of online multiplayer games, AIs now must compete with the experience of playing with other people. To enhance AI behaviors to match that of a real player, the paper discusses the one solution for creating models that can be used for further AI research. Through utilizing some of the built-in features of Unity as well as Photon Network services, the game Maze Escape combines the multiplayer aspect of FPS games and some simple game AI models to allow them to be compared against each other in order to more easily recreate multiplayer experience using AI bots. Thus, this paper hopes to encourage developers to think about how AIs are not only used to enhance single player experiences, but it can also be used in multiplayer.

KEYWORDS

Multiplayer FPS Game, 3D Modeling, AI Machine Learning.

1. INTRODUCTION

As the U.S enters into the 21st century, the sources of entertainment ranging from music, sports, and movies have all been expanding in support of consumerism culture [4]. One of the major industries that have prospered in this time of competition are video game companies. Due to the global pandemic, more people have been spending their time indoors and subsequently video game industry giants such as Riot Games, Blizzard, and Epic games have all tried to produce new games for its audiences in order to maximize their profit [5]. Connected via gaming platforms, the video game industry is reaching its peak of prosperity generating explosive numbers of concurrent players and breaking historical records; the scariest thing being that this trend does not seem to be slowing down anytime soon.

One of the most popular categories of video games that have been played by millions of people around the world are "FPS"s, or First-Person-Shooter, which are popular among those who enjoy the combat experience of playing in the 3-dimensional world through the eyes of the character model [6][7]. While it is included in part of the sub-genre of shooter games, developers are free to shift away from the traditional path that focuses solely on combat with a set of firearms and specific objectives to achieve. They can incorporate their own ideas into the game. Throughout the development of "Maze Escape", the game explores the wide range of tools that are available

for developing AIs in video games in an attempt to select the best solutions for recreating multiplayer experiences.

Some of the existing AI techniques and systems that have been proposed include Navmesh for AI pathfinding, and reactive systems such as finite state machines and behavior trees [8]. Navmesh technique is constructing a mesh from which AI is able to determine the best path from getting from one point to another as well as providing a way to “see” the environment around the AI. Reactive systems allow the AI to interact with the player’s decisions and take different actions to create different experiences for the player. Finite state machines (FSMs) achieve this by having different states the AI can be in to match different existing scenarios in the game while Behavior tree archives this by having a set of rules and conditions to guide the decision making process [10]. These systems have allowed developers to make more complex AI for their games and focus on crafting experiences that are more interactive. Developers have the ability to create AI experiences equivalent to that of other players. However, a bad version of AI can often lead to AI taking away the overall enjoyment of the game rather than adding to it. [Example of a game that has weak AI compared to their multiplayer system]. Thus, developers need good tools to allow them to make good AI.

For instance, Unity provides a built-in nav-mesh system for pathfinding [9]. Pathfinding is the idea in which the player is able to determine the ideal traveling route to take during the navigational processes that takes the least amount of time between two points. The Unity Asset stores would be another case in point. It contains many reactive system tools like FSMs and Behavior Trees. For example, Candice AI has components for handling player detection and combat systems for the game AI. Another example would be Breadcrumbs AI . There are also non-character based AI like Procedural Level Generator, which creates the level around the player themselves.

Our goal is to create a game in which AI can match up to the experience of multiplayer. The current methods that are currently available in the game include Navmesh systems and way point systems. First, there are some good features of the Navmesh system which allows for pathfinding inside the level generator which prevents the player from walking straight through walls and objects in the environment. Second, some good features of the way point system is that it allows for the escaper bot to have an unpredictable route for reaching the end goal which would imitate that of a real player style of playing the game; filled with randomness and does not conform to a fixed way of playing the game. The way point systems generate different routes for the escaper AI bot to take to reach the end goal. When the game starts the way point system will randomly choose a path for the AI to take out of the many the developers had encoded into the system. Thus, we believe that using these simple tools we can at least get close to emulate multiplayer experience in a single player environment. Our system is a synthesis of the Navmesh system and the way point system allowing them to work together for pathfinding while other tools keep them separated.

In two application scenarios, we demonstrate how the above combination of techniques increases the accuracy or the imitation for the ability of the AI to mimic the multiplayer experience.

First, we demonstrate the usefulness of our approach by a comprehensive case study on the competitive AI compared to the competitive player in which the player has to play against one other opponent either AI or another player in a 1v1 game mode of Maze Escape in which one act as the “Defender” and the other as the “Escaper”. Second, we try to analyze the result of our 1v1 competitive AI to a competitive player with a cooperative AI compared to a cooperative player. In this case, not only are we testing the player’s experience playing against an opponent AI and seeing the differences between an actual person and the AI, we are also implementing a way for

the player to play with an ally in the game either AI or another player against two opponents in a 2v2 format. In scenarios, we can test our results through a questionnaire by asking the players' experiences playing the game with the AI versus that of an actual person. Therefore, by measuring the level of artificial intelligence imitation that has been achieved, we are able to verify whether or not the performance of each of the AIs, both the competitive version and cooperative, have been enhanced in order to emulate the real online multiplayer experience. Through the development of these complex, sophisticated and futuristic AIs, certainly, these goals are only one step further and will be explored later in the paper.

The rest of the structure of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample while Section 3 goes into the details of our solutions corresponding to the challenges that we mentioned in Section 2. Next, Section 4 presents the relevant details about the experiment we did which includes other similar experiments done towards this same topic, followed by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Deal with networking

To start off, the first challenge that we encountered has to deal with networking. For instance, in order to create the game, we need to first figure out which Unity Networking solution to use, which is difficult as creating custom networking scripts is much more difficult and challenging compared to using a pre-built solution that requires no coding and programming. Next was setting up the player. We had trouble determining the location of the spawn points and had issues with the player not spawning in the right locations. Also there were errors in the camera fov in which the character model's eyes will see the interior of the gun that blocks off its vision and also the controls as well when the players spawn and not be controlling its own character. There were also errors in the movement and shooting as well. Lastly, setting up the teams required the creation of the teams and in order to keep each side balanced and assign bots accordingly or the game would not start.

2.2. Deal with the navmesh system

Second challenge had to deal with the entirety of the navmesh system. First, we are required to find how to create the navmesh on top of a level. It was difficult to dynamically generate the navmesh during the game as when generating a new level, the nav mesh would not change and the problem with that is the inability to produce an accurate path that matches the level produced by the procedural maze generator that we implemented at the beginning of the game. We also looked for the Nav Mesh component package to make generating nav mesh easier as we kept the nav mesh as a single mesh to allow easy movement from one part of the level to another. We had trouble getting the bots to use the newly produced nav mesh as we couldn't figure out how to get the AI to "walk" with the player model and using NavMesh agent for walking as it required a separate animator component for AI movement to allow the bot to play the walking animation. Lastly, we had to program the code for the escaper bot's waypoint list and randomly choose one point to go to before heading to the end goal in which we designed to allow randomness for the path chosen just like that of a real player.

2.3. Figure the AI's behaviors

Lastly, figuring the AI's behaviors and goals to get it to emulate a real player was the most difficult among the rest. There was an issue with the escaper bots not moving to the end goal and also choosing the same path despite including the random way points that were designed to prevent fixed AI behavior from occurring. Next, there was also a movement issue with the defender bots as they were unresponsive and remained standing throughout the duration of the game and we had difficulty adding the "wandering" system to have them move around the level. The last two issues were for the escapers not trying to run away from the defender when it opened to its field of view and it decided to just run past it and lastly was with the bots not shooting when the opponent player was facing them as it entered its shooting radius.

3. SOLUTION

To begin, the overall system starts with the menu section in which includes the ability for the user to login with a name, then join a game room or create their own, choose their own team. After the initial process, the host of the game room will start the game once every user is ready. The host will also check if any kinds of bots in the game are added. Next, Maze Escape is a strategic multiplayer-FPS game in which players are deployed into a virtual 3D world and are divided into two teams, the Defenders and the Escapers. Because of its competitive nature, in order for both to achieve victory, they will need to satisfy their winning conditions. For instance, the Defender needs to either eliminate all Escapers or keep them occupied until the timer hits zero in order to win. On the other hand, the Escapers are able to either kill all Defenders or try to avoid them and escape the Maze to achieve victory.

We used Unity as the main Game Engine to develop this game and because the is a multiplayer FPS, we also used Photon to manage the multiplayer functionality specifically managing the network connections between the user's computer while syncing all the changes such as player movement, health bar, damage output/input from one client's device to every users' screen thus enabling the multiplayer experience. After this, we are able to implement AI Navigation mesh and way point system for pathfinding and later adding shooting mechanics, animation for the AIs, and AI healthbar in order to recreate the experience of that of a real person.

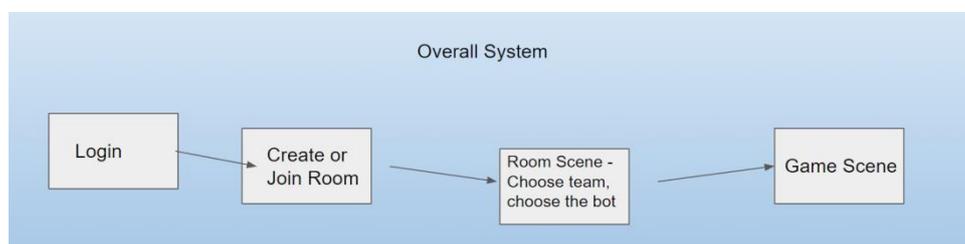


Figure 1. Overview of the system

To start off with, the first component that we implemented was Networking using the photon networking unity version 2.0 package. We built the subcomponents for the menu and the main level for the game. In the menu system, we implemented a sign in system that allows the user to join the game with a gamer tag, a room selection screen that allows the user to decide which active game to join or to create one themselves, and a game room scene in which displays the number of players, robots and their chosen roles. Furthermore, in the game level, we implemented networking by adding specific components to synchronize player position, health, and any animations that are playing. The game level has a component for handling the player and

bot spawning which has a list of points where each team can spawn in as well as a randomizer function to allow positions to be randomized. We also added a game manager system for synchronizing the time left for the round and the amount of players remaining. Based on the win/lose condition the game manager would track which team has won and which team has lost which is shown on the player's screen accordingly. In addition, we have added a script for handling the end goal for the escapers which is a simple trigger which will determine whether the escaper has entered the escaping zone and if true, it will display on the screen for all the players that the escapers have won.

```

1  using System.Collections;
2  using System.Collections.Generic;
3  using UnityEngine;
4  using Photon.Pun;
5
6  public class NetworkedShooting : MonoBehaviourPun
7  {
8      [SerializeField] Camera playerCamera;
9      public float damage = 10;
10     public KeyCode shootKey = KeyCode.Mouse0;
11     public bool canFire = false;
12     private void Start() {
13         if(playerCamera == null)
14         {
15             playerCamera = Camera.main;
16         }
17     }
18     public void Shoot(RaycastHit RCHit, int dmg=0)
19     {
20         if(!photonView.IsMine || playerCamera == null) return;
21         RCHit.collider.GetComponent<NetworkedHealth>()?.TakeDamage(dmg==0?damage:dmg);
22     }
23 }

```

Figure 2. Screenshot of code

Now moving on to the AI behaviors section, we started off with using Unity's built in navmesh system for pathfinding. There are two sides to the navmesh systems. First, the navmesh level component will generate all the pathways from one point in the level to another. Second, the navmesh agent, the AI pathfinder, which uses the generated mesh to determine the optimal path from one point in the level to another. With these two components, the escaper bots are able to find a path to reach the end goal because the navmesh agent will choose the optimal path which is usually the shortest path, the escaper bots tend to always choose the same path resulting in predictable behaviors. To remedy this issue, the waypoint system is administered to each escaper bot with a random waypoint that the AI has to go to before reaching the end goal. This also allowed the defender bots to randomly choose a waypoint to patrol that area. With the pathfinding system mostly updated, to make the AIs more like a player, we added the animation that is similar to the players animations as well a field of view system to allow the bots to react to other players within the level.

4. EXPERIMENT

4.1. Experiment 1

In order to verify that our solution can effectively solve problems at different levels and have good user feedback, we decided to select multiple experimental groups and comparison groups for several experiments. For the first experiment, we want to prove that our solution works stable and continuously, so we choose a group size of 100 different trials in 5 different kinds of target. The 5 different types of skin are cars, stationary enemy, moving enemy, animals, buildings. The goal of the first experiment is to verify if the AI auto targeting algorithm works good for different types of targets. Through sampling 5 groups of different targets. Result is collected by statistics if the AI could find the target correctly. Experiments have shown that almost all targets in different

types tested the right result. Moving enemy has the most correct rates, which means our user are works more better in aiming the moving enemy in the game. This experiment could explain that the data sets do have a obvious impact on the finding the targets and aim it, because the data we are using have a high rates of the moving enemy. The average correct rate of 5 different types of the aims shows below:

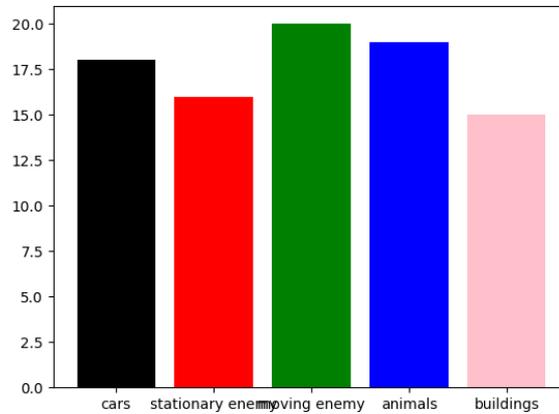


Figure 3. Result of experiment 1

A good user experience is as important as a good product. So a perfect solution should have excellent user experience feedback. In order to prove that our solution has the best user feedback, we specially designed a user experience questionnaire base on the US system usability questionnaire rules. We statistics the feedback result from 100 users, Track the user's data for 10 days play, let them explore freely on the functionality of the game. We divide those users into Five different groups. The first group of users ages from 10 - 20, the second group of users ages from 20 - 30, the third group of users ages from 30 - 40, the fourth group of users ages from 40 - 50, the fifth group of users ages from 50 - 60. The goal of the first experiment is to verify high feedback scores show high performance. We collect the feedback scores form these 5 different groups of users and analyze it. Experiments have shown that users who ages from 20 - 30 give the highest result feedback to our game. Which may because of the age between those range are more likely to play a shooting game and using the auto targeting scheme The experiment graph shows below:

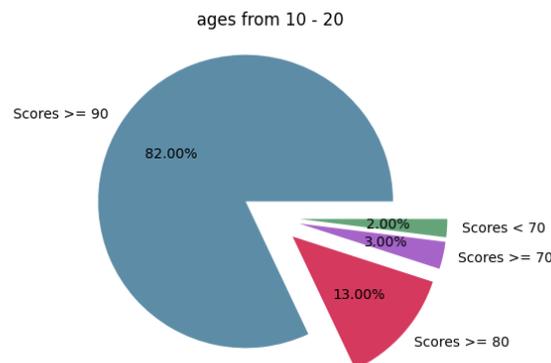


Figure 4. Result of age 10-20

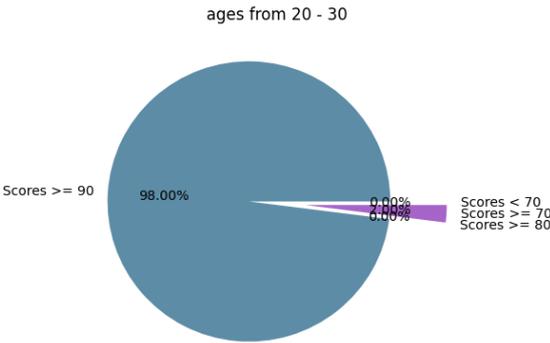


Figure 5. Result of age 20-30

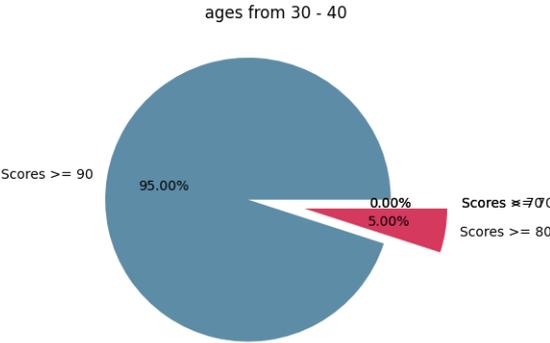


Figure 6. Result of age 30-40

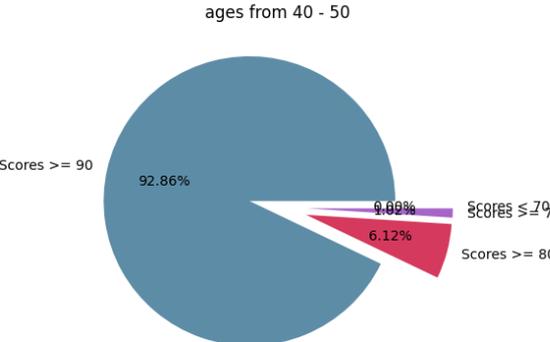


Figure 7. Result of age 40-50

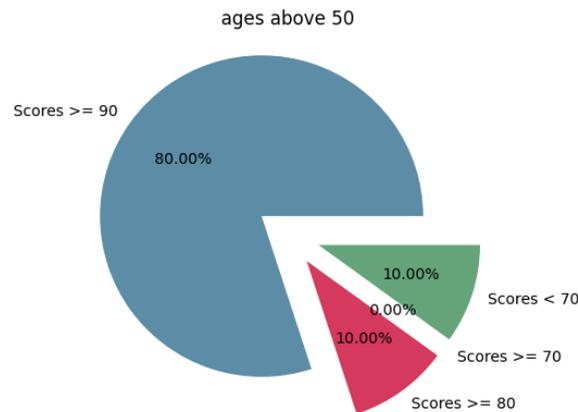


Figure 8. Result of age above 50

5. RELATED WORK

Omid Shafieid et al describes how games have contributed to the research of artificial intelligence [11]. When decomposing the technical nature of games, AIs can absorb the information in order to learn other complex tasks which is similar to our goal as we continue to pursue the enhancement of AI and super-human performance level to imitate that of a real player. For instance, this work uses graphical analysis to identify how each scenario is solved through a list of fixed sets of policies which are designed to encapsulate diversified interactions within the game which is extremely helpful as it helps generate creative game structures that can be used to train AIs.

Livingstone describes how the Turing's test can define the association of human behavior with ai through a series of interrogation questions and how successfully an AI is able to mimic human responses under certain circumstances [12]. This is similar to our works as we both try to explore the level in which the imitation of human-like responses can be generated by a machine. Some differences however is that the Turing test requires a human operator that engages in a series of questioning while our game's AI specifically focuses on the imitation of human-like behaviors.

Barata at el describes how it was difficult for a massive multiplayer strategy game that has a rapidly changing environment which requires the players to be active for a long time in order to encourage realism within the game [13]. They implemented a MMOTBSG as a replacement for human players right when they become inactive is a strength of their research as the AI plays the game for players with similar level of performance which is quite impressive. This also parallels our research as we seek to increase AI behavior to match that of a real player in order to emulate the multiplayer fps experience.

6. CONCLUSIONS

Maze Escape is a game that combines multiplayer with the AI experience. In order to enhance the multiplayer experience, we can replicate the behaviors from that of a real player onto the AI through using a list of methodologies which enables the AI to have human functions that is distinct from that of the AI experience. Maze Escape is excellent in determining the optimal solution for combining the multiplayer experience with the AI as it provides the option for the user to choose the AI to either be the opponent who enables the competitive aspect of the game or

choose the AI as a cooperative partner which enables the cooperative aspect of the game. All of which is essential for recreating the multiplayer experience as it fulfills its goal of having the AI imitate the human player. Maze Escape is unique as the escaper and defender AIs all have its own goal of reaching the winning condition and in order to achieve that, they will be able to mimic the responsibilities from that of a real player in order to satisfy the requirements for the game to end and ultimately for one side to achieve the ultimate victory. When developing AIs, others should consider the level in which the bots are able to provide an experience which will match that of a real player just as if the bots were human players not only for single player games but as well as multiplayer games.

Since the game is still in the developing phase, the AI behavior is limited to two main states as in the current AI it is only able to go to someplace and react if it sees an opponent which they will be programmed to either shoot or run away. The game itself is very simple because it only has one role, and one map level so it is not able to sustain the interplay of an environment that contains multiple variables/interactions between the game and the player.

In the future, the game will be adding an additional role system for players to choose their unique class division which contains special weapons and abilities that are designed just for that class. There will also be updates to maps, and animations specifically for different abilities based on player roles. With respect to the AI updates, there will be addition to AI states to allow more actions for the AI and different model types such as decision trees, behavior trees, and planner systems.

REFERENCES

- [1] Cillessen, Antonius HN, and Peter EL Marks. "Conceptualizing and measuring popularity." *Popularity in the peer system* (2011): 25-56.
- [2] Regensburger, Alois, et al. "Photon propagation in a discrete fiber network: An interplay of coherence and losses." *Physical review letters* 107.23 (2011): 233902.
- [3] Corchado, J. M., and B. Lees. "Integration ai models." *Workshop On Knowledge Discovery And Data Mining*. Pml-Nerc, Plymouthlondon, UK. 1998.
- [4] Hanif, Hanif, and Is Susanto. "Consumerism Culture Of Urban Communities Based On Islamic Economic Perspective." *AGREGAT: Jurnal Ekonomi Dan Bisnis* 4.1 (2020): 83-99.
- [5] Mofijur, Md, et al. "Impact of COVID-19 on the social, economic, environmental and energy domains: Lessons learnt from a global pandemic." *Sustainable production and consumption* 26 (2021): 343-359.
- [6] Jansz, Jeroen, and Martin Tanis. "Appeal of playing online first person shooter games." *Cyberpsychology & behavior* 10.1 (2007): 133-136.
- [7] Cardamone, Luigi, et al. "Evolving interesting maps for a first person shooter." *European Conference on the Applications of Evolutionary Computation*. Springer, Berlin, Heidelberg, 2011.
- [8] Brewer, Daniel. "Tactical pathfinding on a navmesh." *Game AI Pro 360: Guide to Tactics and Strategy* (2019): 25-32.
- [9] Lester, Patrick. "A* pathfinding for beginners." [online]. *GameDev WebSite*. <http://www.gamedev.net/reference/articles/article2003.asp> (Acesso em 08/02/2009) (2005).
- [10] Brand, Daniel, and Pitro Zafiropulo. "On communicating finite-state machines." *Journal of the ACM (JACM)* 30.2 (1983): 323-342.
- [11] Omidshafiei, Shayegan, et al. "Navigating the landscape of multiplayer games." *Nature communications* 11.1 (2020): 1-17.
- [12] Livingstone, Daniel. "Turing's test and believable AI in games." *Computers in Entertainment (CIE)* 4.1 (2006): 6-es.
- [13] Barata, Alexandre Miguel, Pedro Alexandre Santos, and Rui Prada. "AI for massive multiplayer online strategy games." *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*. 2011.

- [14] Harrison, Fiona E., et al. "Spatial and nonspatial escape strategies in the Barnes maze." *Learning & memory* 13.6 (2006): 809-819.
- [15] Cutumisu, Maria, and Duane Szafron. "An architecture for game behavior ai: Behavior multi-queues." *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. Vol. 5. No. 1. 2009.

© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

COMPARISON OF FORECASTING METHODS OF HOUSE ELECTRICITY CONSUMPTION FOR HONDA SMART HOME

Farshad Ahmadi Asl¹ and Mehmet Bodur²

¹Mathematics and Computer Science Department, Faculty of Arts and Sciences,
Eastern Mediterranean University, Famagusta, via Mersin 10, Turkey

²Computer Engineering Department, Faculty of Engineering, Eastern
Mediterranean University, Famagusta, via Mersin 10, Turkey

ABSTRACT

The electricity consumption of buildings composes a major part of the city's energy consumption. Electricity consumption forecasting enables the development of home energy management systems, resulting in the future design of more sustainable houses and a decrease in total energy consumption. Energy performance in buildings is influenced by many factors, like ambient temperature, humidity, and a variety of electrical devices. Therefore, multivariate prediction methods are preferred rather than univariate. The Honda Smart Home US data set was selected to compare three methods for minimizing forecasting errors, MAE and RMSE: Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Fuzzy Rule-Based Systems (FRBS) for Regression by constructing many models for each method on a multivariate data set in different time-terms. The comparison shows that SVR is a superior method over the alternatives.

KEYWORDS

Forecasting, Mathematical Models, Electricity, Prediction, Consumption, ANN, SVR, FRBS.

1. INTRODUCTION

Honda Smart Home was constructed in California, USA, with the goal of creating a sustainable home and a zero-carbon lifestyle [1]. In the domestic energy sector, the development of optimization methods such as *maximum power point tracking* made the use of energy sources like photovoltaic solar energy economically feasible decades ago [2]. Recent studies indicate that the economic optimization of renewable energy in domestic energy consumption can be further extended by enhancing power management. According to studies, buildings are responsible for the largest proportion of energy consumption in a city, and the residential section is a significant part of it [3][4]. Heating, Ventilating, Air Conditioning (HVAC) systems, and lighting are the main energy-consuming sources of domestic houses. Domestic energy consumption has been increasing due to several factors, like globalization, greenhouse gas emissions, and population growth [5][6]. For the same reasons, the importance of increasing energy efficiency grows, and electricity forecasting plays a key role in it [7]. For a variety of applications, including management, optimization, and energy conservation, the importance of accurately forecasting the energy consumption of buildings is emphasized [6]. In addition, accurate energy forecasting models have many implications for the planning and energy optimization of buildings and are crucial to the economy [8]-[10]. Consequently, energy management utilizing optimized energy

forecasting of a domestic house can increase the energy efficiency of the house and the utility of renewable sources such as solar panels.

In the remaining parts of this text, Section 2 explains the time terms and data preparation; Section 3 discusses the details of the models; Section 4 describes the evaluation performance of the models, and Section 5 concludes this work.

2. THE HONDA SMART HOME DATA SET AND TIME TERMS

2.1. The Honda Smart Home Energy Management System Dataset

In the Honda Smart Home project, which began in 2015 and is ongoing, data related to energy management, such as the HVAC system and Home Energy Management Systems (HEMS), are recorded at a one-minute sampling rate. This forecasting performance study is based on six-month-long energy consumption data from October 2020 to March 2021 of the Honda Smart Home project. The data set is sparse because many of the electric devices work on or off by the resident's decision.

2.2. Forecasting Time Terms and Data Preparation

Nonlinear-Multivariate Machine Learning (ML) models for domestic electricity consumption forecasting are built and compared with each other in three-time terms. *Medium-term electricity load forecasting* (MTELF), usually for a week up to a year, which is useful for maintenance scheduling and planning power system outages[11]. *Short-term electricity load forecasting* (STELF), for intervals ranging from one hour to one week; and, one of its primary applications in the daily operation of the electric power system [12]. *Very short-term electricity load forecasting* (VSTELF) ranges from a few minutes to an hour ahead, which is applicable for real-time control, as practiced by [13].

The VSTELF of this study used ten random samples of seventy-minute data collected over a six-month period to get statistical parameters for forecasting performance evaluation. Similarly, the STELF and MTELF used ten pieces of four-day and two-month data randomly selected within the six-month data period for performance comparison of the models, assuming that randomly selected windows reveal the model's weaknesses and strengths better by covering different modes of power consumption.

The original data set contains attributes with one-minute sampling intervals. The VSTELF and STELF followed the original data set's one-minute sampling interval. The MTELF is resampled every ten minutes to reduce the length of the data set to a reasonable size for the forecasting process.

The data set is subdivided as shown in Table 1 in order to evaluate the performance of forecasting independently.

All data sets are prepared in a matrix with eight numerical scalar input variables (attributes) and one scalar target output variable. But occasionally, when a random data set has a column or columns of zero values, some of the forecasting models are unable to scale the data to predict the output value. The problem can be solved by omitting the column(s) containing zeros. This circumstance only occurs in VSTELF due to the small size of the data set. Therefore, the input attributes of this time term are variable.; therefore, the input attributes of this time term are variable.

Table 1. Data sets obtained from the Honda database.

<i>Models</i>	<i>Data sets</i>	<i>Sampling Periods (minute)</i>	<i>Size of Data set</i>	<i>Time Covered</i>	<i>No. input Attributes</i>
BRNN	VSTELF training + verification	1	30 + 30	1 hr.	Variable
SVR & SBC	VSTELF training	1	60	1 hr.	Variable
All models	VSTELF test	1	10	10 mins	Variable
BRNN	STELF training + verification	1	1500 + 1500	2 days	8
SVR & SBC	STELF training	1	3000	2 days	8
All models	STELF test	1	3000	2 days	8
BRNN	MTELF training	10	2000 + 2000	1 month	8
SVR & SBC	MTELF training	10	4000	1 month	8
All models	MTELF test	10	4000	1 month	8

The input variables are the measurement of outdoor temperature and air humidity, as these are significant external factors that affect the house's electricity consumption; and the average power consumption of electric devices (lighting of the living room, lighting of the kitchen, washing machine, refrigerator, microwave, and fans), as these are components of the majority of houses' power consumption. The sum of the average power consumption of the mentioned devices is the target output.

Each attribute in the data matrix has its own units and data range. For successful forecasting, each attribute and output variable are normalized by linearly mapping the columns of the data matrix to the interval [0, 1].

3. FORECASTING MODELS

This study tested a number of forecasting models based on the three mentioned methods, which are employed successfully by researchers in regression analysis studies. The best model for each method in terms of performance accuracy was chosen for comparison and evaluation: Bidirectional Recurrent Neural Networks (BRNN) [14]; Support Vector Regression with Analysis of Variance Radial Basis kernel Function (ANOVA RBF) [15][16]; and a combination of the Subtractive Clustering (SBC) method and the Fuzzy C-Means [17][18].

3.1. Forecasting by Bidirectional Recurrent Neural Networks

The first model is BRNN, which is computationally expensive compared to basic ANN models such as feed-forward neural networks (FFNN), but the results have shown that the forecasting performance of this model is more accurate for this case study. A basic FFNN and a set of RNNs, including Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM), with both ReLU and Tanh activation functions, were tested. The error was calculated by taking the average of 10

runs, and the BRNN with Tanh showed marginally superior performance accuracy with the 10 randomly selected data sets [19]. Figure 1 compares basic FFNN and BRNN models for 2 days of forecasting.

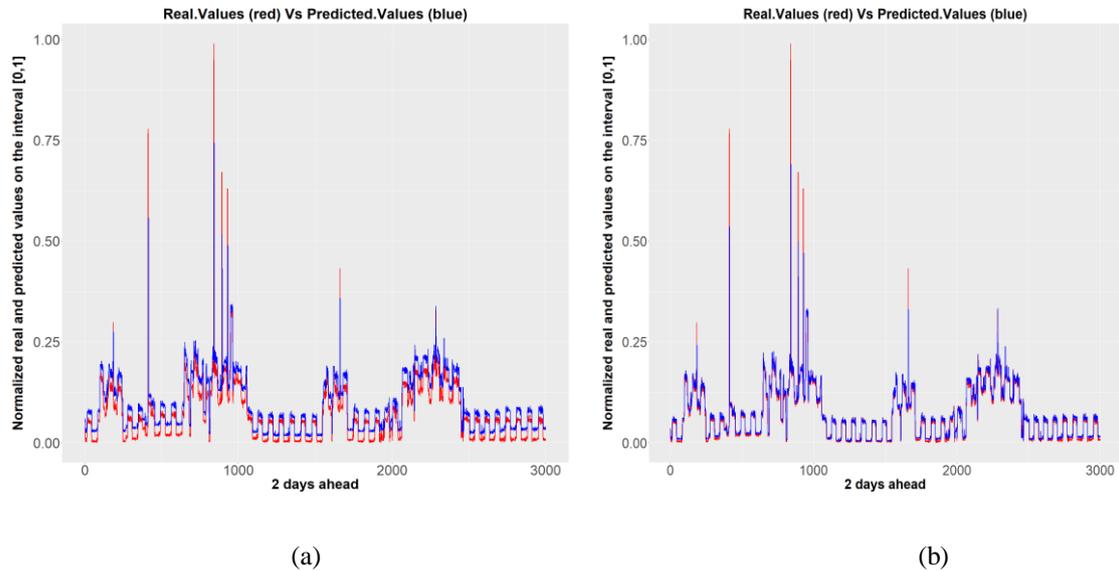


Figure 1. Plots of (a) basic FFNN and (b) BRNN models for 2 days forecasting.

3.2. Forecasting by Support Vector Regression Method

The second model, SVR with the ANOVA RBF kernel, is tested from the KERNLAB package [20]. Following a search of various packages and kernel functions such as RBF, Tanh, Bessel, and Laplace, finally, ANOVA RBF from the KERNLAB library package provided outstanding forecasting performance. The SVR method supports regression tasks and employs the Sequential Minimal Optimization (SMO) algorithm. SMO reduces execution time by breaking down the search into multiple sub-search tasks. The SVM saves computational effort by managing the

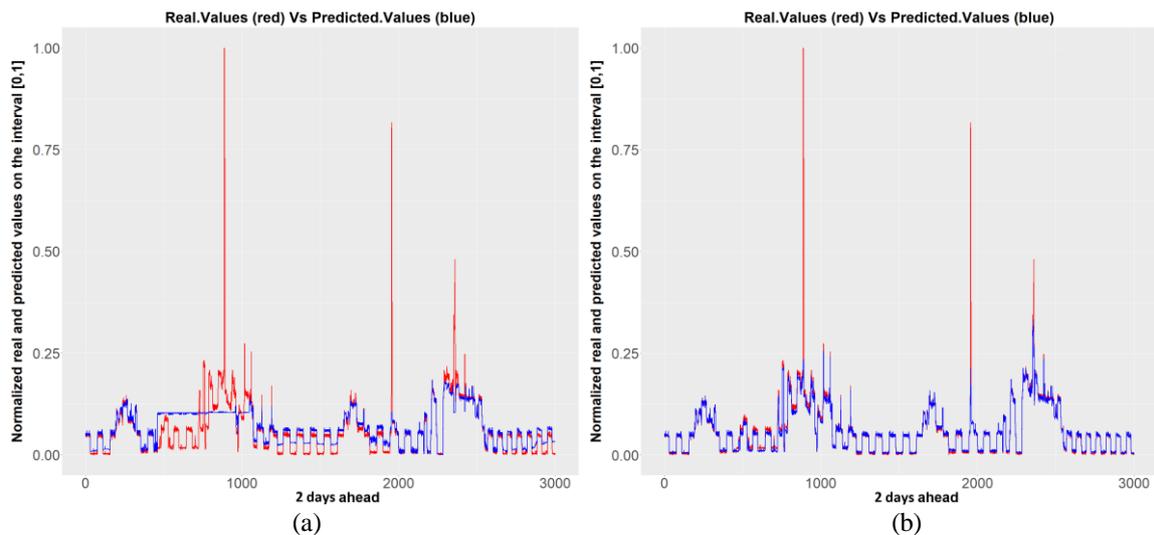


Figure 2. Plots of the SVR with the (a) RBF and (b) ANOVA RBF Kernels.

fitting process and the modelling process simultaneously [20]. Figure 2 illustrates the plots of the SVR models with the RBF and ANOVA RBF kernels next to each other from the KERNLAB library package for a better comparison.

3.3. Forecasting by Fuzzy Rule Base System with Subtractive Clustering

The third model is a combination of the SBC method and the Fuzzy C-Means technique from the FRBS package [22]. SBC considers each data point as a potential cluster centre and calculates the likelihood of each data point defining a cluster centre based on its distance to all other data points. The point with the highest potential among the remaining points is chosen as the next cluster centre. Afterward, the process repeats until all cluster centres are obtained. The Fuzzy C-Means algorithm is then used to optimize the cluster centres [17].

A set of models such as Fuzzy Rule-Based Systems based on space partition, neural networks, clustering approach, and the gradient descent method were evaluated from the FRBS package to determine the model with the highest performance accuracy. The majority of them resulted in poor forecasting performance and long runtimes, which made them inadequate for forecasting very short-term electricity consumption. Consequently, the SBC model has been considered acceptable for power consumption forecasting. Side-by-side comparisons of the Hybrid Neural Fuzzy Inference System (HyFIS) and the SBC model with C-mean optimization are presented in Figure 3. Both models are available in the FRBS package.

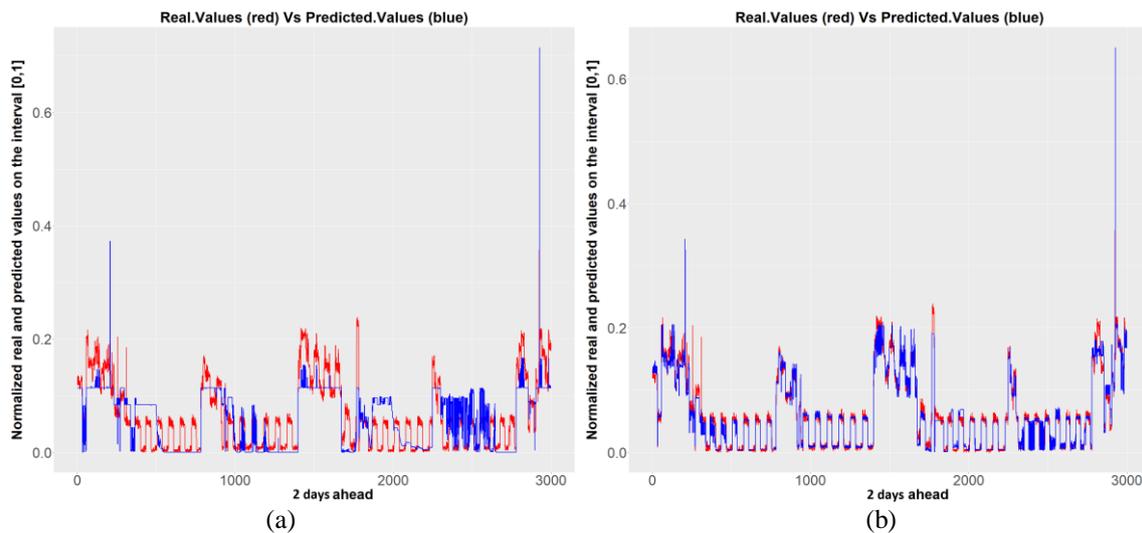


Figure 3. Plots of the models, (a) HyFIS and (b) SBC.

All of the written codes for the models in the R programming language can be found in [23].

4. MODEL PERFORMANCE EVALUATION

The accuracy of all models is measured in two metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These two metrics are used in this work as they are the most common for measuring the accuracy of electricity forecasting and thus make this study more comparable to the others. The models are comparable since they use the same test data and normalization method ([0–1] min–max normalization). Additionally, the models' execution times are measured. Table 2 contains the average MAE and RMSE values of 10 repeated runs of each model for each time term. The average execution time of the 10 repeated runs of each model for each time term is given in Table 3.

Table 2. Evaluation of the models: Average of 10 runs with 10 random data samples.

<i>Time-Terms</i> <i>Models</i>	<i>VSTELF</i>		<i>STELF</i>		<i>MTELF</i>	
	<i>MAE</i>	<i>RMSE</i>	<i>MAE</i>	<i>RMSE</i>	<i>MAE</i>	<i>RMSE</i>
<i>SVR</i>	0.009	0.013	0.005	0.019	0.0038	0.014
<i>SBC</i>	0.027	0.042	0.01	0.027	0.0055	0.014
<i>BRNN</i>	0.11	0.126	0.011	0.031	0.024	0.046

Table 3. Execution Time of the models: Average of 10 runs with 10 random data samples.

<i>Time-Terms</i> <i>Models</i>	<i>VSTELF</i> <i>Execution time</i>	<i>STELF</i> <i>Execution time</i>	<i>MTELF</i> <i>Execution time</i>
	<i>SVR</i>	less than 10 sec	less than 10 sec
<i>SBC</i>	less than 10 sec	Around 5 min	Around 9 min
<i>BRNN</i>	Around 15 sec	Around 30 sec	Around 45 sec

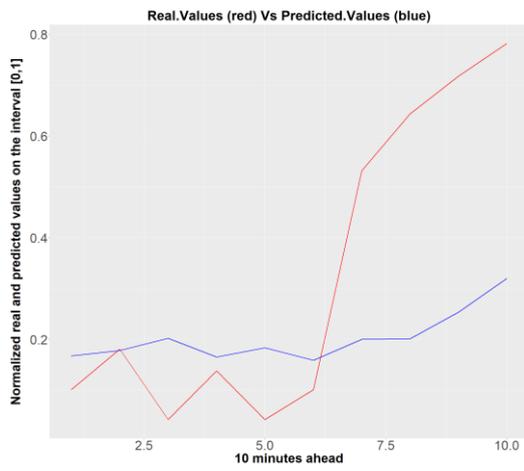
As demonstrated in Figure 4, the BRNN and SBC models are not ideal for very short-term forecasting. In contrast, the SVR model is suitable for VSTELF with a decent result. The execution times of the models are around 10 seconds in this time term.

The performances of the SBC and BRNN models for short-term forecasting are fairly similar. The results showed that these two models are more accurate with a larger data training set, whereas the SBC model's execution time becomes noticeably longer as the training set grows. The performance and execution time of the SVR model compared to the other two models are better in this time term. Model plots for side-by-side comparison are shown in Figure 5.

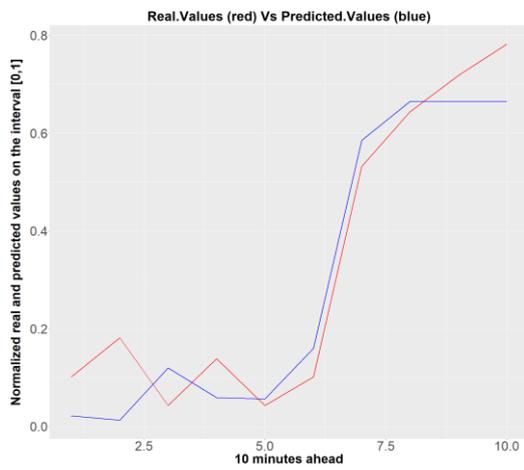
The BRNN model performs marginally worse than the SBC for medium-term forecasting. It can be the result of data set sampling. On the contrary, SBC showed its best forecasting performance. SVR is the model with the most accurate forecasting performance among the others. The execution times of the models are nearly identical to STELF, with the exception of SBC, which requires approximately four minutes longer to complete. Figure 6 compares the plots of the models in this time term.

Similar to the research that showing STELF is a suitable area for the implementation of neural networks, the BRNN model in this study showed its best performance in this time term [24]. The SBC model demonstrated decent performance in terms of performance accuracy while working with a large data training set, but it makes it inadequate for VSTELF. Also, it was observed that the SBC model could not predict all of the output values and express them as undefined values.

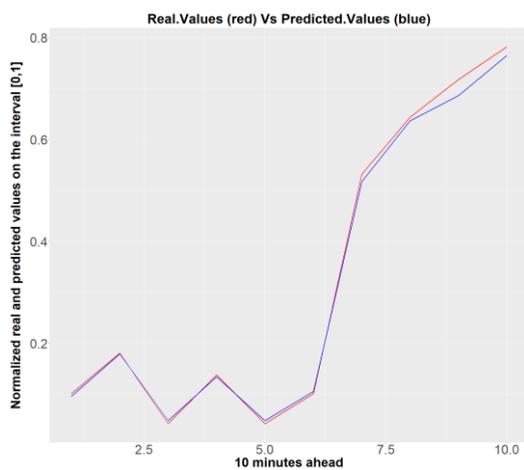
Both SVR and SBC models were unable to scale the model to predict the output value when a random data training set contained zero values in one or more columns. The issue was solved by removing the zero column/columns from the data.



(a)

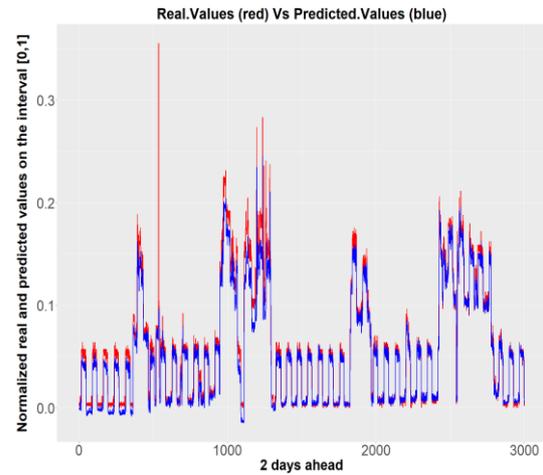


(b)

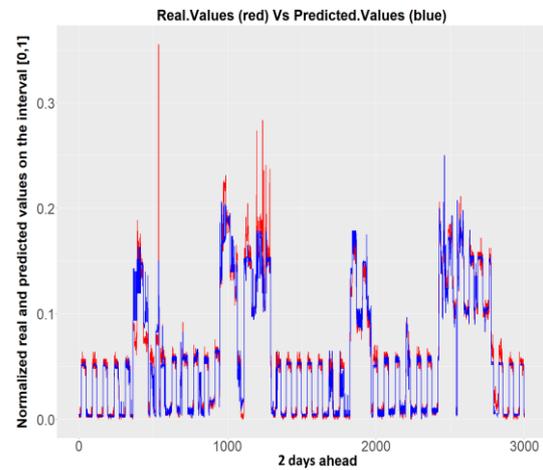


(c)

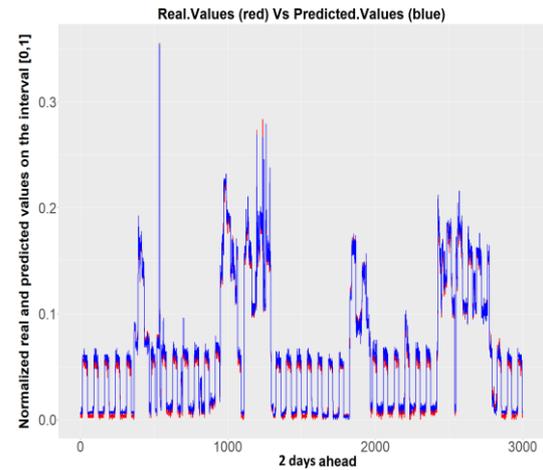
Figure 4. Plots for (a) BRNN model, (b) SBC model, (c) SVR model in VSTELF.



(a)



(b)



(c)

Figure 5. Plots for (a) BRNN model, (b) SBC model, (c) SVR model in STELF.

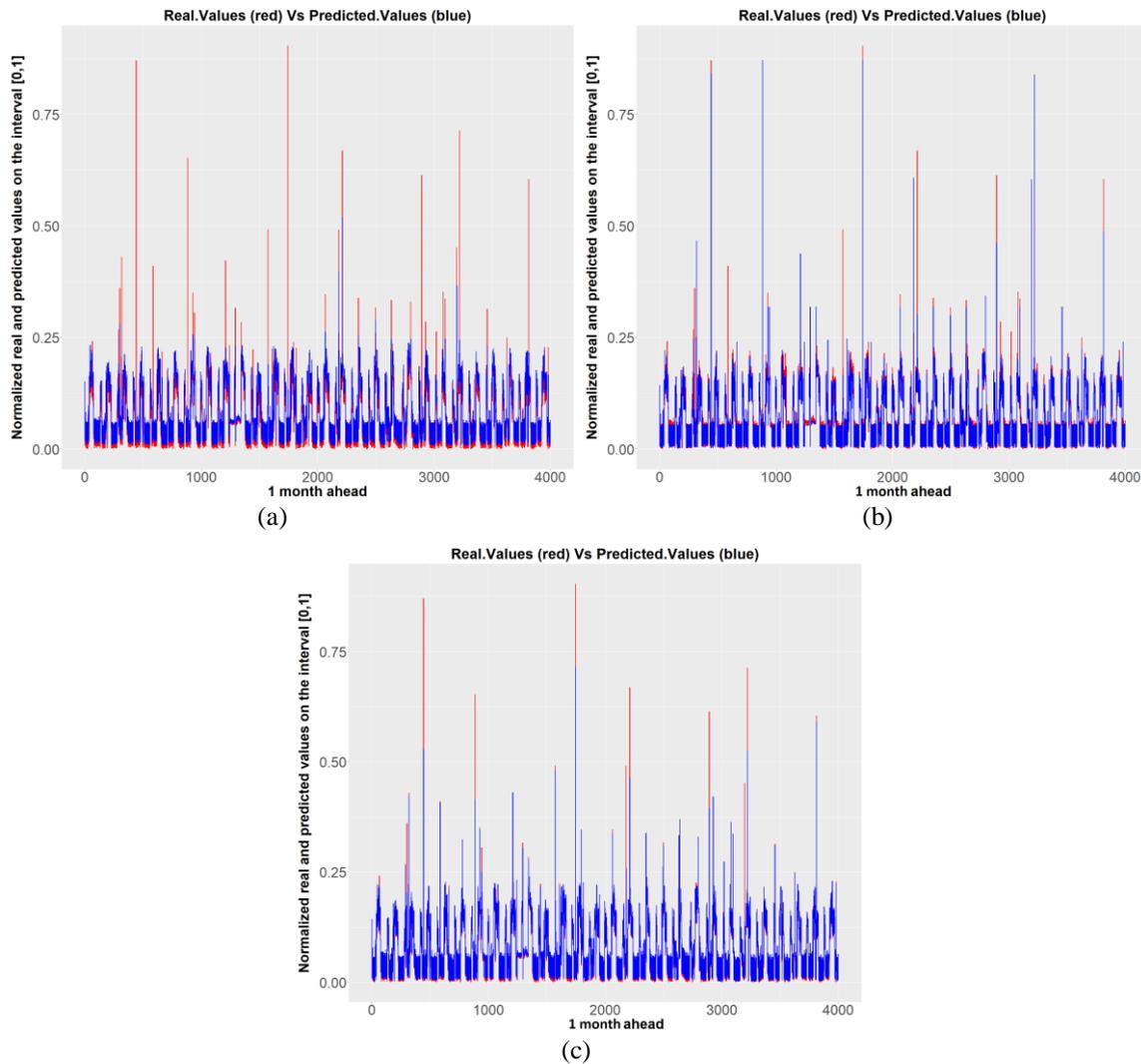


Figure 6. Plots for (a) BRNN model, (b) SBC model, (c) SVR model in MTELF.

5. CONCLUSIONS

This study compared the performance of three widely-used methods for forecasting the electricity consumption of domestic houses over three forecasting time terms. The strengths and weaknesses of each method were observed across different data set sizes, time terms, and execution times. In addition, the best model constructed using a single method is identified for this case study.

The BRNN model uses a bidirectional layer that processes a sequence in both directions, making the model ideal for time series forecasting [19]. This is one of the reasons why this model is more accurate than the other RNN models tested. In addition, the Tanh activation function activates almost all the input neurons to predict the output, which makes it more computationally expensive but more accurate than the other activation functions, thereby enhancing the model's performance accuracy. Nevertheless, this model is suitable for STELF and MTELF but not ideal for VSTELF.

The SVR model was constructed using the KERNLAB library package and employs an SMO optimization algorithm during the modelling process, which is significantly faster than data

deduplication techniques such as the chunking algorithm on sparse data sets [21]. ANOVA, by analysing and comparing differences between group means or population means (of variables) and their associated procedures, such as variation [25][26], helps the RBF kernel and the model for a precise forecast. This model handles both the fitting and the modelling processes at the same time, saving computational effort and making it suitable for forecasting in all time terms.

In terms of forecasting accuracy, the SBC model with the Fuzzy C-mean optimization outperformed the other tested models in the FRBS library package. However, it is inefficient for very short-term forecasting, especially with a large training data set, but it is suitable for STELF and MTELF if a long execution time is not a concern for the forecast.

The results indicate that the selected SVR model forecasts with a lower mean absolute and root mean square error than the other models in all time terms. Additionally, this model is suitable for very short-term forecasting since its execution time is fast, even for large data sets. Moreover, the simple implementation of the SVR model makes it an excellent choice for forecasting in all time terms for time series data.

REFERENCES

- [1] Honda Smart Home Us, (2021) Retrieved from International Living Future Institute: <https://living-future.org/lbc/case-studies/honda-smart-home-us>.
- [2] M. Bodur and M. Ermis, (1994), "Maximum power point tracking for low power photovoltaic solar panels," *Proceedings of MELECON '94. Mediterranean Electrotechnical Conference*, pp. 758-761 vol.2, DOI: 10.1109/MELCON.1994.380992.
- [3] Aurelie Foucquier, Sylvain Robert, Frederic Suard, Louis Stephan, Arnaud Jay, (2013) State of the art in building modeling and energy performances prediction: A review *Renewable and Sustainable Energy Reviews* 23, 272–288, DOI: 10.1016/j.rser.2013.03.004.
- [4] Khosravani, Hamid R., Castilla, María Del Mar, Berenguel, Manuel, Ruano, Antonio E. & Ferreira, Pedro M. (2016) A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building, *Energies*, 9(1), 57; 9(1), 57, DOI: 10.3390/en9010057.
- [5] Bot K., Ruano A. & da Graça Ruano M. (2020) Forecasting Electricity Consumption in Residential Buildings for Home Energy Management Systems. In: Lesot MJ. et al. (eds) *Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU. Communications in Computer and Information Science*, vol 1237. Springer, Cham; DOI: 10.1007/978-3-030-50146-4_24.
- [6] Runge, J.; Zmeureanu, R. (2019) Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. *Energies*, 12, 3254, DOI: 10.3390/en12173254.
- [7] Mathieu Bourdeau, Xiao Qiang Zhai, Elyes Nefzaoui, Xiaofeng Guo, Patrice Chatellier, (2019) Modeling and forecasting building energy consumption: A review of data-driven Techniques. *Sustainable Cities and Society* 48, 101533, DOI: 10.1016/j.scs.2019.101533.
- [8] Deb, Chirag, Zhang, Fan, Yang, Junjing, Lee, Siew Eang & Shah, Kwok Wei. (2017) A review on time series forecasting techniques for building energy consumption, *Renewable and Sustainable Energy Reviews*, Volume 74; Pages 902-924, DOI: 10.1016/j.rser.2017.02.085.
- [9] S. J. Kiartzis, A. G. Bakirtzis, J. B. Theocharis, and G. Tsagas, (2000) "A fuzzy expert system for peak load forecasting. Application to the Greek power system," *10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No.00CH37099)*, 2000, pp. 1097-1100 vol.3, DOI: 10.1109/MELCON.2000.879726.
- [10] G. Gross and F. D. Galiana, (1987) "Short-term load forecasting," in *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1558-1573, Dec. DOI: 10.1109/PROC.1987.13927.
- [11] Abu-Shikhah, Nazih & Elkarmi, Fawwaz. Medium-term electric load forecasting using singular value decomposition. *Fuel and Energy Abstracts*. 36. 4259-4271, DOI: 10.1016/j.energy.2011.04.017.
- [12] Friedrich, Luiz & Afshari, Afshin. Short-term Forecasting of the Abu Dhabi Electricity Load Using Multiple Weather Variables. *Energy Procedia*. 75. 3014-3026, DOI: 10.1016/j.egypro.2015.07.616.

- [13] Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020) Methods and Models for Electric Load Forecasting: A Comprehensive Review, *Logistics & Sustainable Transport*; 11(1), 51-76, DOI: 10.2478/jlst-2020-0004.
- [14] M. Schuster and K. K. Paliwal. (1997) "Bidirectional recurrent neural networks," in *IEEE Transactions on Signal Processing*; vol. 45, no. 11, pp. 2673-2681, Nov., DOI: 10.1109/78.650093.
- [15] Drucker, Harris and Burges, Christopher J. C. and Kaufman, Linda and Smola, Alex and Vapnik, Vladimir, (1996) Support Vector Regression Machines, MIT Press.
- [16] Fisher, R. (1954) The Analysis of Variance with Various Binomial Transformations. *Biometrics*, 10 (1), DOI: 10.2307/3001667
- [17] S. Chiu, (1996) "Method and software for extracting fuzzy classification rules by subtractive clustering", Fuzzy Information Processing Society, NAFIPS; pp. 461 – 465.
- [18] James C. Bezdek, Robert Ehrlich, William Full, (1984) FCM. The fuzzy c-means clustering algorithm, *Computers & Geosciences*; Volume 10, Issues 2–3, Pages 191-203, ISSN 0098-3004, DOI: 10.1016/0098 3004(84)90020-7.
- [19] Chollet & Allaire. (2017). RStudio AI Blog: Time Series Forecasting with Recurrent Neural Networks. Retrieved from <https://blogs.rstudio.com/tensorflow/posts/2017-12-20-time-series-forecasting-with-recurrent-neural-networks>.
- [20] Karatzoglou A, Smola A, Hornik K, Zeileis A. (2004) "kernlab – An S4 Package for Kernel Methods in R." *Journal of Statistical Software*; 11(9), 1–20.
- [21] J. Platt. (2000) Probabilistic outputs for support vector machines and comparison to regularized likelihood Methods, *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schoelkopf and D. Schuurmans, Eds. Cambridge, MA: MIT Press; DOI: 10.1.1.41.1639.
- [22] Riza LS, Bergmeir C, Herrera F, Benitez JM. (2015) "FRBS: Fuzzy Rule-Based Systems for Classification and Regression in R." *Journal of Statistical Software*; 65(6), 1–30.
- [23] <https://github.com/FFarshadd/Electricity-Forecasting>
- [24] A. T. Sapeluk, C. S. Ozveren, and A. P. Birch, (1994) "Short term electric load forecast using artificial neural networks," *Proceedings of MELECON '94. Mediterranean Electrotechnical Conference*, pp. 905-908 vol.3, DOI: 10.1109/MELCON.1994.380955.
- [25] Horst Langer, Susanna Falsaperla, Conny Hammer, (2020) Chapter 3 - Unsupervised learning, Editor(s): Horst Langer, Susanna Falsaperla, Conny Hammer, In *Computational Geophysics, Advantages and Pitfalls of Pattern Recognition*, Elsevier, Volume 3, Pages 8124, ISSN 2468547X, ISBN 9780128118429, DOI: 10.1016/B978-0-12-811842-9.00003-0.
- [26] Damien Chanal, Nadia Yousfi Steiner, Raffaele Petrone, Didier Champagne, Marie-Cécile Péra, (2021) Online Diagnosis of PEM Fuel Cell by Fuzzy C-Means Clustering, Reference Module in Earth Systems and Environmental Sciences, Elsevier, ISBN 9780124095489, DOI: 10.1016/B978-0-12-819723-3.00099-8.

A SOCIAL-DRIVEN INTELLIGENT SYSTEM TO ASSIST THE CLASSIFICATION OF PET EMOTIONS USING DEEP LEARNING AND BIG DATA ANALYSIS

Hans Li¹ and Yu Sun²

¹Damien High School, 2280 Damien Ave, La Verne, CA 91750

²California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

Pets have always been a big part of families, and people always imagine what their pet is thinking by their actions and face [1]. However, No one can tell what a pet might be thinking unless they are very familiar with them [2]. This paper develops/designs/proposes an application/software/tool to... [There is an AI that could try to understand the pet's emotion, and people can share their photos to other pet lovers on the app, which will further make the AI more accurate [3]. We applied our application to people who have pets of all kind and conducted a qualitative evaluation of the approach. The results show that [the AI is decently accurate and the app is fairly easy to use from feed backs made by testers. And the AI have the potential to get more accurate in the future with the more data customers posts, and thus will give more accurate results back to the users].

KEYWORDS

Social-Driven, Machine Learning, Classification.

1. INTRODUCTION

Pets are a very important part of our family, and pets are very popular in recent years [10]. People always wonder what their pet might be thinking, or want to share their pet's funny moment with other people, specifically people who have the same feeling— pet owners. That's why I built an app for this society, and with AI, which with everyone's help, can be more accurate in determining what their pet is thinking, and get their beloved pet what they might need or want.

Some of the apps have similar features, however, they mainly focus on the AI part, they can trace the face and come up with a expression, but that's about it, just a tool [4]. It's not very accurate, and you can't do anything after getting the result. And the app is not well designed to store those pictures, which they probably don't have a database for.

I have the share feature or post feature which can let you share your pet's movements every time you take a picture or do an expression check. My method is inspired by all the social apps, but they are mostly focusing on people and stuff, and is a more general app. There are some cool features like, posting after you got an result from the AI, saving your pets and organizing the posts, and in a server that everyone can see it. We believe that every pet lover wants to share

about the fun moments of their pets to someone who really knows about pets, or that is also a pet lover.

After training the AI for a period of time, I introduced the app to my friends. 4 out of 7 of my friends say it is somewhat accurate to their knowledge, and 2 of them thinks the AI is not accurate on their pets. Afterward, 2 sets of 10 tries, and the accuracy was 7/10 and 9/10. the app feature is in general fairly smooth to use. with not a lot of pages, the app isn't complicated and confusing. After some testing from people around me, I received good accuracy feedback on average.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Using New Language

While using this app, it is my first time using this language, and it is very confusing or frustrating sometimes when you know what you want to write but you don't know what's the word for a function, or you get confused with another language and type a whole lot wrong. But after you get used to it, it would not be as hard anymore since the logic is similar in every language.

2.2. How to design

Another challenging part is how to design, this is my first time doing any app or website, And i found it hard to achieve what i want the pages to look like, It is all very simple and boring shapes the the code can offer, and it takes a long time to create any interesting design i want [5]. It is also a challenge to think of what the users may like.

2.3. Loading speed

The final challenge I faced on this project is it is a big project for the emulator, which I use to test the app on, and it just takes very long to load [6]. I ended up using a real android phone instead by connecting it to my laptop, and it works pretty smoothly with the developer mode. But I also found out it looked a bit different on the real phone compared to the emulator, and that is mainly because of the screen size difference. But this also reminded me of the different sizes of screens so I can make improvements to my design.

3. SOLUTION

The app works like any other social media apps [7]. You have to log in first [8]. You can log in with an email and password. If you don't have an account yet, you can register one with your email, and it will ask you to create a password. Then your account data will be safely stored at Firebase's server. It is all connected so no worry about the account getting stolen. And when you are in the homepage, there will be a post button, which you can post whatever you wanna post,

word, picture, and your result from your AI. And the AI will be in the post page if you need. There will also be a pet's page, which is all about your pet, your page, which is about you [9]. And a navigation bar at the bottom to get you to places fast.. The AI is trained, and the more picture we get from the app post, we will further train the AI more. Which will make it more accurate. overtime. the AI sees the picture and will try to figure out all important points of the picture and compare it with the known-picture and answers to give the final result.

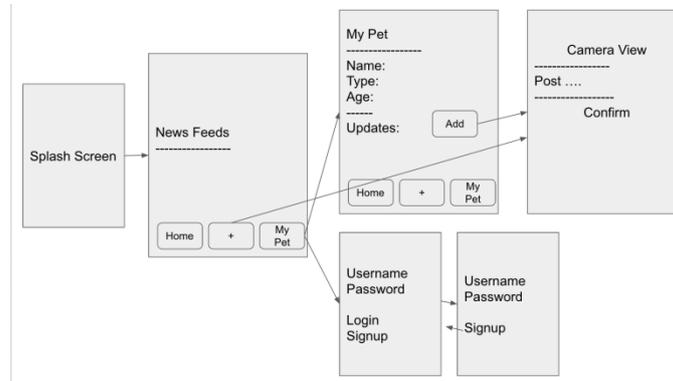


Figure 1. Overview of the solution

```

void LogIntoFireBase() async {
  if(submitLock) {
    return;
  }
  submitLock = true;
  try {
    UserCredential userCredential = await
    FirebaseAuth.instance.signInWithEmailAndPassword(email: email, password: password);

    //global.user;

    submitLock = false;
    Navigator.pushReplacement(
      context,
      MaterialPageRoute(builder: (context) => HomePage()),
    );
  }
}

```

Figure 2. Code of Login

This part above is the login page of my app, these specific lines are for whether the login email and password is true or not. It will get the email and password, then it will send it to Firebase, firebase will give an answer back whether it found it or not, if so, the lock will turn false and the user will be let in, if not then it will still be locked to this page.

```

}on FirebaseAuthException catch(e){
    if(e.code == 'user-not-found') {
        popUpInfo("No user found");
    }
    else if(e.code == 'wrong-password') {
        popUpInfo("Wrong password");
    }
    else{
        popUpInfo("Unknown error");
    }
    submitLock = false;
}
}

void popUpInfo(String message) async{
    ScaffoldMessenger.of(context).removeCurrentSnackBar();
    ScaffoldMessenger.of(context).showSnackBar(SnackBar(content: Text(message)));
}

```

Figure 3. Code of Firebase

This part of code is to let the user know why their login failed. Firebase will receive the email and password, then if it didn't match up with the database, it will send back the reason, and my app will tell the user in a pop up banner about why it failed to login. Is it the username or the password that is incorrect. Or, it might also be unknown if Firebase didn't send anything back.

```

class NavigationBar extends StatelessWidget{

    void navigateToHomePage(BuildContext context){
        Navigator.pushReplacement(
            context,
            MaterialPageRoute(builder: (context) => HomePage()),
        );
    }

    void navigateToPage2(BuildContext context){
        Navigator.pushReplacement(
            context,
            MaterialPageRoute(builder: (context) => Page2()),
        );
    }

    void navigateToUserProfile(BuildContext context){
        Navigator.pushReplacement(
            context,
            MaterialPageRoute(builder: (context) => UserPage()),
        );
    }
}

```

Figure 4. Code of Navigation Bar

This is the navigation bar, which will be at the bottom of the homepage. it is where the buttons are and where you can click and visit the other pages, the code above it for where the button will send the users to, which is to homepage, user page and post page.

```

bool validForm(){
  if(email.length == 0){
    popUpInfo("Please enter your email.");
    return false;
  }
  if(Nickname.length == 0) {
    popUpInfo("Please enter your Nickname.");
    return false;
  }
  if(password.length == 0){
    popUpInfo("Please enter your password.");
    return false;
  }
  if(password != confirmPassword) {
    popUpInfo("Your passwords does not match.");
    return false;
  }
  return true;
}

void popUpInfo(String message) async{
  ScaffoldMessenger.of(context).removeCurrentSnackBar();
  ScaffoldMessenger.of(context).showSnackBar(SnackBar(content: Text(message)));
}

```

Figure 5. Code of registration page

This is the registration page, you have to input an username, an email, and password, and confirm the password. If the password doesn't match, then it won't be able to pass.

```

-
} on FirebaseAuthException catch(e) {

  if (e.code == 'weak-password') {
    popUpInfo('Error: The password provided is too weak.');
```

Figure 6. Code of Firebase Auth Exception

This part of the code is all the reason why the registration won't pass, it depends on what Firebase gets back to the app. you can't use an already-used email, or an invalid email. And it could fail if your password is too weak.

```

Widget PetPageBody(Map<String, dynamic> petData){
  String uid = FirebaseAuth.instance.currentUser.uid;

  bool isowner = petData["ownerID"] == uid;

  print("Pet Page");
  print(uid);
  print(petData["ownerID"]);
  print(isowner);

  Column c = new Column(
    children: [
      Text('Pet Name: ${petData["name"]}'),
      Text('Birth Year: ${petData["birthYear"]}'),
      Text('Animal Type: ${petData["species"]}'),
      Text('Desc: ${petData["bio"]}'),
      Text("TODO SHOW PET'S POSTS"),

      Expanded(child: SingleChildScrollView(child: PetPostList(widget.petID)),

      ElevatedButton(onPressed: (){ navigateToUserProfile(petData["ownerID"]); },
child: Text("Owner Page")),

    ]
  );
}

```

Figure 7. Code of profile page

This is in the pet profile page, it is a page where the pet information the user enters before hand will stay, the app will go to Firebase, ask for the pet under the user's id, and print out all the information it have. which includes name, birth year, species and a brief description from the owner, the user can use the pet information to make a quick post when they choose the pet they are posting on the pet post page.

4. EXPERIMENT

4.1. Experiment 1

I tried a couple of pictures to test the AI's accuracy. I have 10 different pictures, and I see that it gets the same result every time. The ideal result is every set I try, every picture comes back with the same result, which proves that the machine isn't guessing but really identifying it.

Try	Results
1	happy
2	happy
3	happy
4	happy
5	happy
6	scared
7	scared
8	happy
9	angry
10	happy

Figure 8. Result of experiment 1

The result came back decently, 7 out of 10 pictures in every set came back the same, which is good enough for a new AI machine, because it will come better with more training.

4.2. Experiment 2

The second experiment I did is I introduced my code to a couple of people around me, and I let them taste their own pictures of their pet, and compare it to what they think their pet's emotion is in their own opinion.

people	result	expectation	feedback
1	sad	depressed	positive
2	sad	boring	positive
3	happy	nothing	positive
4	happy	angry	negative
5	angry	scared	positive
6	sad	happy	negative
7	happy	cozy	---

Figure 9. Result of experiment 2

4 out of 7 friends gave me positive feedback on the AI machine, and 2 came back negative, which means I can still work on and train my AI better later on.

The AI machine meets my expectations. , and from the feedback, i made some improvements on it. My expectation is it is a stable AI, which it did decently in, and it is at least 50% right on the determination of the result.

5. RELATED WORK

Happy pets is an app available in the app store, it's focus is on the AI part, it can find out the animal, the breed, the emotion from a picture [11]. and it also has a description of every breed of cats and dogs built in. Mine is different in a way that my AI is not capable of finding some of the stats like "happy pets" does, but i have a social feature.

Pettitude is an app available in the app store, it's only focuses on the emotion on the pet, and that's also the only feature of it [12]. Comparing to mine, it has the similar AI feature, but without the social feature that my app is including. and doesn't store any data as well. But it doesn't require internet to use.

My friend has an app that could tell the bark of a dog [13]. It is similar in some ways to the AI, but in general, it is for a different use than mine. Mine is for social and his is for security. But the AI both plays a significant role in the app.

6. CONCLUSIONS

I have created a app that is dedicated to pets and pet owners to share their story and with the tools to make it fun, such as AI [14]. The app is meant to share your pet and spread happiness among people, and ask questions for people who have the same hobby or interest as you. The app contains every essential thing for a social app. And we store information on Firebase. It requires a login to use, and people can input their pet's information on their pet page. The AI is another thing important to this app. The AI is fairly accurate and will get better as more pictures and comments are stored by the app. It can read the picture and give an estimated result of the emotion of the pet from its facial expression [15]. The experiment results show it is in a good stage. A high percentage of people said it was accurate. And it is very stable. Which is important. So I would say it is solving the question I had at first place-" I want to know what my pet is thinking and share it to other pet lovers".

Even though it is evolving, the accuracy is still far from scientifically accurate. The app it sells could also look better and perform better with more advanced codes. And I can improve the controls and looks of the app after customer's surveys or likes.

I will create an AI from the start, and try to train it to my best. I can read multiple kinds of animals clearly. It can be more accurate. And i will ask or search more on designing, and improve my overall app quality.

REFERENCES

- [1] Archer, John. "Why do people love their pets?." *Evolution and Human behavior* 18.4 (1997): 237-259.
- [2] Paul, Elizabeth S., et al. "Sociality motivation and anthropomorphic thinking about pets." *Anthrozoös* 27.4 (2014): 499-512.
- [3] Bao, Katherine Jacobs, and George Schreer. "Pets and happiness: Examining the association between pet ownership and wellbeing." *Anthrozoös* 29.2 (2016): 283-296.
- [4] Ekman, Paul. "Facial expression and emotion." *American psychologist* 48.4 (1993): 384.
- [5] Gemperle, Francine, et al. "Design for wearability." *digest of papers. Second international symposium on wearable computers (cat. No. 98EX215). IEEE, 1998.*
- [6] Ahrenholz, Jeff, et al. "CORE: A real-time network emulator." *MILCOM 2008-2008 IEEE Military Communications Conference. IEEE, 2008.*
- [7] Kang, Xiaoyu, et al. "Delivery of instructions via mobile social media app increases quality of bowel preparation." *Clinical Gastroenterology and Hepatology* 14.3 (2016): 429-435.

- [8] Purdy, George B. "A high security log-in procedure." *Communications of the ACM* 17.8 (1974): 442-445.
- [9] Allen, Karen. "Are pets a healthy pleasure? The influence of pets on blood pressure." *Current directions in psychological science* 12.6 (2003): 236-239.
- [10] Cohen, Susan Phillips. "Can pets function as family members?." *Western Journal of Nursing Research* 24.6 (2002): 621-638.
- [11] IRAWAN, DEBBIE. *Analysis and Design of Website and Web-based Internal Management System for Happy Pets Company*. Diss. BINUS, 2007.
- [12] Wu, Jie, and Yunsheng Wang. "Social feature-based multi-path routing in delay tolerant networks." *2012 Proceedings IEEE INFOCOM*. IEEE, 2012.
- [13] Lupovici, Amir. "The Dog That Did Not Bark, the Dog That Did Bark, and the Dog That Should Have Barked: A Methodology for Cyber Deterrence Research." *International Studies Review* 23.4 (2021): 1672-1698.
- [14] van der Linden, Dirk, et al. "Pets without PETs: on pet owners' under-estimation of privacy concerns in pet wearables." *Proc. Priv. Enhancing Technol.* 2020.1 (2020): 143-164.
- [15] Tian, Ying-Li, Takeo Kanade, and Jeffrey F. Cohn. "Facial expression analysis." *Handbook of face recognition*. Springer, New York, NY, 2005. 247-275.

MASS SURVEILLANCE, BEHAVIOURAL CONTROL, AND PSYCHOLOGICAL COERCION THE MORAL ETHICAL RISKS IN COMMERCIAL DEVICES

Yang Pachankis

Universal Life Church, Modesto, California, USA

ABSTRACT

The research observed, in parallel and comparatively, a surveillance state's use of communication & cyber networks with satellite applications for power political & realpolitik purposes, in contrast to the outer space security & legit scientific purpose driven cybernetics. The research adopted a psychoanalytic & psychosocial method of observation in the organizational behaviors of the surveillance state, and a theoretical physics, astrochemical, & cosmological feedback method in the contrast group of cybernetics. Military sociology and multilateral movements were adopted in the diagnostic studies & research on cybersecurity, and cross-channeling in communications were detected during the research. The paper addresses several key points of technicalities in security & privacy breach, from personal devices to ontological networks and satellite applications - notably telecommunication service providers & carriers with differentiated spectrum. The paper discusses key moral ethical risks posed in the mal-adaptations in commercial devices that can corrupt democracy in subtle ways but in a mass scale. The research adopted an analytical linguistics approach with linguistic history in unjailing from the artificial intelligence empowered pancomputationalism approach of the heterogenous dictatorial semantic network, and the astronomical & cosmological research in information theory implies that noncomputable processes are the only defense strategy for the new technology-driven pancomputationalism developments.`

KEYWORDS

Cybersecurity, Risk Prevention, Psychosocial Cybernetics, Cyber Surveillance, Time & Entropy, Human Trafficking, Defense on Outer Space, Defense Strategy, Decision Theory, Sexism in LGBTQIA+ Rights, Satellite Information Paradigm.

1. INTRODUCTION

In solving the problem of cyber surveillance with commercialized personal devices, I adopted a scientific practice with the FITS convention on outer space research, informed by multilateralism. [1] Since my original theory has no precedent protocols in any established frameworks, my empirical research started with the interactive features of the JS9-4L online software after adopting a semantic analytical theory. [2] [3] [4] [5] [6] Traditional paradigm on human security depends on the Big Bang cosmology, but the maladaptation of the holographic principles to the control theory on utilizing globalization to dictatorial nationalism pose the fundamental threat to global peace & security. [7] [8] In counteracting the malicious militant threats with various forms of cyber-based coercion by dictatorial semantics, I asserted a democratization thesis for the sake of psychodynamic induction. [9] [10] The methodology in the research overtly counteracted the

covert military operations of the People's Liberation Army by the limitation of spiders needed in ontological networks and organizational structures. [11] Self-efficacy and mental counter-measurements adopted the psychological techniques introduced by Daniel Kahneman in relation to the Cognitive-Affective-Behavioural model. [12] [13] The panic attacks were mainly regulated by psychological detachment from the communication networks. [14] [15] The dual use of rationality & psychology is diagnostic for the technician blindspots caused by professionalism with a pancomputationalism belief. The rationality in quantum information & informatics, with dissected domains, assisted in the preservation of mental health in absolute rationalism regarding the trivial satellite applications of dictatorial command chains. [16] It is with the non-psychological self I present the discussions in the moral ethical dimensionality that has intruded to the level of commercialized personal devices. Since there are numerous methods in targeted trafficking including in commercial settings of advertisements, the technicalities I would like to address here is specifically on counteracting the geostationary orbits that are manifested for various breaches in humanitarian laws & human rights by covert military actions.

2. METHODOLOGY

The methodology is consisted of four parts. The first part of the methodology provides an analytical framework on the underlying physics principles in satellite applications. It is supplemented by the information theory on shell efficacy in the second part. The developmental science method also serves the purpose in growth-mindset building in resilience to the multilateral dynamics. Since topology is key to the ontological frameworks and artificial intelligence in immoral & unethical satellite applications with totalitarian big data practices, the third part proposes a DNS-free ontology as the theoretical part of the methodology in the stead of the traditional paradigm on human security. [11] [7] Concerning the satellite kernels are bound for efficacy reasons that artificial intelligence might be used in inter-satellite linking for GUI monitoring, the topological constructs in isolated domains are theorized in relation to quantum gravity, which produced the scientific evidence with tangent vectors in relation to astronomical instrumentation. [17] [18] I will explicate the sexism-based psychodynamic observational approach in the counter-measurement of the dehumanizing cyber space of the P.R.C. that I had been using for the preservation of my gender identity, which enabled my research process.

2.1. Analytical Framework

Geostationary orbits follow either Newtonian gravity in time protocols (TP) or general relativity (GR). The principles of GR governs the communication capacities of information transmittance. Unlike the observational space-based telescopes that can adapt atomic or light clocks for time management beyond the aforementioned physical principles in terms of gravity, predatory satellite networks are bound to organizational principles of certain human groupings. Current security constructs depend on I/O & cryptography, and the surveillance networks' operations depends on geostationary communication across topological domains in the earthbound physical signals - added with an extra layer of server-side command I/O chain. [19] This means that security breaches to the cybernetics occur at the bottom-most domains in any topological constructs, even though the human organizations & commands of the surveillance networks are bounded by Newtonian principles. The intrusive surveillance technologies not only breach privacy data, but also propriety data in forming new mirror images & mounts. Therefore, the spheres of physical signals have been the central concern and points in breaching by the adversaries.

Handshake protocols seek to resolve the issue with time series in data points, but the cost is proportional to the inefficacy of the method. Moreover, handshake protocols do not resolve

satellite attacks and ping attacks, i.e. denial of service (DoS) attacks with consecutive time series. This fundamentally breaches the traditional paradigm of human security and financial systems. [7] Such attacks are institutionally manifested by regime approaches to multilateralism, disguised in culture with heterogeneous cryptography, and parasite on globalization with surrogacy economy & human surrogacy in various sectors, regardless of being coerced, in unawareness, or being active accomplice. [10] [9] The transactional cryptocurrency by centralized banking is only one contemporary military use of the method for marginal gain in the market economies. [21] Privacy breaches in this regard, is only a collateral damage of such cyber espionage. Disconnection of the server data to the user interfaces can be a technical & physical solution to the attacks but not available for commercialized personal devices. As long as the interest for targeted surveillance remains, alternative methods can still arise with the criminal organizations. [22]

The truncation of informatics in defense uses is bound for high bias values depending on the adversaries' database. Satellite kernels are effective for personal devices with dynamic responses similar to DNS environment, depending on the relative locations between the device & satellite. [23] However, device hard drive in unsafe topological networks is still bound for information breaches that can be utilized by adversary powers, such as cache analysis, hence disruptions on human security. [19] As randomized I/O with multi-channel access & processing can be associated with the space-based instrumentation access points in physical signals, based on the **Baryonic cores** of the new ARM-based chips, the exogenous energy-efficacy-driven information-access architectures in dictatorial controls largely neglected the **Lagrange points** behind the "randomness" in **time series distributions**. [20] [19] Albeit the dynamic communication & information touchpoint(s) are theoretically unpredictable for computational processes, machine learning with artificial intelligence can still count the touchpoint(s) for imprecise predictions in such topological spaces with artificial intelligence floats. [8] The technological demands on corporate compliances further breach any technological & technical developments in privacy & propriety protection with hardware [in]compatibilities with room for further developments in adversary solutions. Since many APPs, despite of excessive uses of cookies, are manifesting on the short-range inter-device ports for information gathering or chain of controls, shell exclusion from the satellite kernel to end user root system(s) is the best alternative for human security protection, which in turn risks of developing a totalitarian cyber environment by the mass produced and uniformly distributed batch shells. [20] In a constructive course in the spirit of science, the "randomness" in the I/O in time series, with the law of small numbers in relation to the law of large numbers, can provide theoretical insights to the concentrated gravitational fields of the cosmos to the **instrumentation local gravity space(s)** in the solar system, and not the exogenous mass surveillance & human trafficking topologies in geostationary orbit with RAID mirror images. [24]

2.2. Shells in Satellite Applications

The purpose of using shells is to block malicious geostationary-orbit satellite applications. Even though geostationary orbits theoretically have the advantage of automation in device targeting and response frequencies, hence human trafficking & trafficking in persons, shell-scripting can reprogram devices' location responses in relation to the Lagrange points instead of to the SaaS topologies from malicious providers in centralized satellite networks. [25] This means that the shell automation creates a relativistic DNS that is not bound for the ontological definitions in the malicious topological networks - at a price of gross privacy breaches to the LGBTQIA⁺ community with the proxy in my 2019 field trip to New York & New Haven for the purpose of getting rid of the human trafficking & fulfilling my homosexual marriage. [25] In order for such shells to be broken, adversaries need to be able to predict the **Lagrange points** in their server time series, which only has mathematical values in realist terms of defense. By any method of deduction it is impossible for the adversary to develop precision strikes to the instrumentation

without the science or original source data behind it, and by abduction the cost for detection & first response on the breached points of contact is relatively low. The only possibility is anti-satellite weapon attacks that were explicitly forbidden by outer space law, but can be achieved with inductive-adductive methods of weaponry designs. [27] However, precision strikes are the least likely because it would require predictability on the Lagrange points from ground-to-space or space-to-space decks and the predictability for its scientific operations from the geostationary-satellite-driven decision-making.

The satellite kernel generated shells will be effective scientific products for the next-generation outer space explorations. Data preserved in the shell-core can have secured API extensions but not bound for time-series based attacks. The vulnerability of such theory-driven designs is with the package encryption. Even though quantum key distribution has been the focus of outer space competitions, individual based targeted hacks can still happen for chained effects on such breaches in distributed shells - as long as the anticipated gain is worth of the adversaries' decision-making. [28] Since raw data deduction is key to the data structures in wearable devices, a deduced-structure construct in such devices' operating systems can effectively prevent the usability in the breached data and can only function with the appropriate associated device(s) and cloud service(s). [29] It means that log file structures are the key security products in the satellite kernel to device shell pacts. The market force of mass-producible products will naturally marginalize the economic forces behind malicious technology developments. The adversaries' niche competition strategy is proven to be **ecological hazard**. [30] However, cross-domain semantic analysis is still a risk for psychological & psychosocial threats in military sociology & human-relation-based coercions, which underlies the reason for human trafficking & trafficking in person that accompanied me even to and in the United States of America. [25] [29] [26]

As I have hedged my homosexual marriage Gestalt psychodynamics to the structural dictatorship of P.R.C. since my undergraduate studies with the discourse of "ménage à trois", sexism had been the discourse & disguise for me in the social-theatre approach under the surveillance networks. [31] [2] [11] As the human trafficking harms the adversaries inflicted on me, LGBTQIA⁺ sexual conducts in morality & ethics are not even in the concern of the adversary power(s) let alone the legitimate human rights of founding a family. [32] As my inherent dignities, fundamental freedom, and autonomy have been severely transgressed all my life from the criminalization of homosexuality in the 1990s on in P.R.C., the cognitive-affective-behavioral therapy has been effective in counteracting the "flash" RAMs with sexism approaches. [25] [13] [33] However, the continued labor trafficking through the telecommunication networks with probable Newtonian gravitation ontologies remains unsolved. [11] Counter-measurement first-aid was received regarding the continued gross harassments & psychological warfare calculations from the PLA, with fake accounts in gay social media & privacy data of the U.S. & U.K. military personnel's. [34] Therefore, my behaviour online as an independent variable to the structural dictatorial human trafficking command chain became a variable to the structural stigma indicators. [26] [35] I was asked to edit an internal pitch video of the Ministry of Public Security of P.R.C. in 2015 with no mandatory confidential contract that I intended to submit to the American government in my trip for marriage in 2019, and the psychological coercions in the exogenous human trafficking surveillance network were embedded in the narratives with the technical information. Detailed information as evidence is in *Appendix A* of the article, along with the poison pen letter addressed to my birth mother following the USCIS letter on U/T VISA process in *Appendix B*. [36] As my abnormal psychological pragmatism in the online interactions faced by human trafficking and human rights abuses has suggested, my physiological presence in the mass surveillance regime is the threat to outer space security. [2] [15]

2.3. Human Centredness in Topology

Since fundamentally, there is no perfect privacy in technological constructs, human centredness in light of the mass psychological cyber warfare is more critical in human rights protection. [11] The shell-distribution theoretically should follow the humanitarian principles with more focus on macro-technical stability in the cloud server side(s) and minimalist principles in device data storage. [4] As the research, by pragmatism solved the information paradox, the scientific aspect of human centredness in relation to the physical sciences in astronomy & thermonuclear astrophysics is still key to the multilateral competition in human trafficking. [37] It is with this regard, the intercepted manuscript behind the white hole (WH) observation is attached in *Appendix C*. Individualized shells may follow the anthropological traces in astrobiological developments, and some specificity is expressed in the ethnicity of the reviewed material. [38][39] Militarily, such approaches are often suspected by the dictatorial powers with top-down restrictions in market economy with import-export controls in centralized banking system, but not for all countries. Instead of a dynamic DNS in traditional IP paradigm, I suggest a theoretical dynamic topology based on Euclidean mathematical constructs for defense, which can feed the adversaries on the global deductive distributions with their already functioning natural resource satellites in their perceived randomness & data point(s) breaches. [40][41] In their minting civilizational ideology, the organizational criminology can be possibly decreased by their own high costs & risks in electronic warfare, an analysis of which will be attached in *Appendix D*. [10] [39] The data points sampling in astrophysics data systems hence can be consistent with the developmental technology purposes in satellite kernel applications in the cosmological & astrophysical sciences. [23] [16] Hence the time series will only be predictable with the quantitative or qualitative selection of kernel & data points in black hole (BH) singularities, which effectively render technology transfer invalid in cyber in the long run. [42][43]

Since brainwaves are analogical to the pulsating of condensed mass objects in time series, the frequency generation in electronic personal devices in connection & contact points don't necessarily need to be purely mechanical. The concept of dynamic topology in inter-device frequency generation can adapt dynamic instead of standardized amplitude, with latter of which bound to be utilized by Wi-Fi security breaches or power-data transmission transgressions. Racism based intelligence surveillance on my 2019 marriage trip exactly used the latter in following me possibly by the FBI or NSA, with inaccurate CIA intelligence that targeted me as an adversary spy when in fact I am a human trafficking victim trying to do the right thing with appropriate legal procedures. This factor also led to the further degradation on American cyber & national security if not outer space security for being forced to reenter the P.R.C. without travel documents & American documents. In normative right courses, intelligent wearable devices can adapt biological truncation such as blood pressure circulation as one of the indicators for psychological & physiological time-keeping, to the natural ecological environment in real-time astronomical data, whereas such conceptualization has more demanding requirements on cloud data security & big data structuring, underlying the GR & Newtonian gravity relevance with space-based telescopes. [44] [18] The human-centred fluidity largely depends on the backend fluxes of gravitational mass from the natural cosmos, whereby constitutes the human-centred adaptation of GR. Whereas satellites like Star-link concentrates on the stability of ground signal transmissions, such theoretical concept derives from the highly elliptical orbit of the Chandra X-ray space telescope. [9] [10] The dynamic orbit itself shields the satellite data from any tampering attempts.

2.4. Quantum Gravity for Network Decentralization

With the IP/TP locality on earth, the Archimedes geometrical distribution of the cosmos is currently covered by the Event Horizon Telescope (EHT) with ionizing radiation, wherefore the

essentiality on the time of being on earth in terms of solar orbits is a geologically bound fraction of numerical display on the elapsed portions of solar radiation. The integral fraction from satellite data kernel distributed geologically bounded IP to devices or data centres of TP that can be quantized as individual fractions of gravity in terms of gravitational mass. The ratios in fundamental (a) symmetry in cosmology and in BH & WH juxtapose thence becomes of a singularity-based alternative to the Big Bang theory with uncertainties in dark energy & dark matter. [18] [43] Semantic constructs are bound for psychological cyber warfare with the linguistic domains of coercion by the exogenous command chains, whereas pure physical-principle based networks can only be decoded by appropriate scientific approaches. [16] [11] [15] Keyboard tracking technologies may extract texts for machine learning and / or automation, however, the essential aspects of the natural sciences are non-interoperable, with scientific results irrelevant to linguistic alterations. [9] [2] This means that anyhow the discourses and / or diplomatic rhetorics change, the network infrastructure is not bound to any human intervention or fabrication - except for technological advancements that can expand the knowledge in the natural sciences or engineering that better reflects the natural world with resistant materials. [43]

The quantum gravitation construct can either build up on the GR version of the codified feedbacks, or the singularity version of time-independent relativistic singularity. [43] Instead of focusing on hardware limitations such as CPUs, relativistic feedback loops in satellite kernels can essentially replace or at least ease the use of caches. [28] [29] As the light properties of ionizing and non-ionizing radiation are historically proven to preserve the quantum information in data systems, the cache in satellite kernels ought to be for the GUI processing I/O networks. [44] [45] This implies that thermal radiation information can serve for the predictive factor for space instrumentations with satellites included. The feedback loop is the non-computable process involved in the series of research conducted on personal / commercial device under the targeted trafficking. The incompatibility of the data-energy ports where my research device is is the factual basis for the previous results, as in my observation when the *FBI* or *NSA* followed me when I was in the U.S. in 2019, possibly for the gross privacy transgressions caused by the targeted trafficking on me with DoS attacks even after I lost the Chinese carrier SIM card in the trip. Essentially, the EHT constitutes the global distribution for the prototype of BH surface information in quantum gravity with global data sampling points, whereby the maritime ambition driven mercantilist approach in globalization by the dictatorial power in high energy LASERs in the case of P.R.C. essentially, with the unethical & irresponsible use of technologies, counteracted the cosmic electron bursts shown in *Figure 1*. In the electroweak regime, polyion applications can be the first steps in the satellite kernel network designs in light of quantum computers with electromagnetic amplitude by energy consumption for the large-scale structure and non-optical order of the universe. [46] [47]

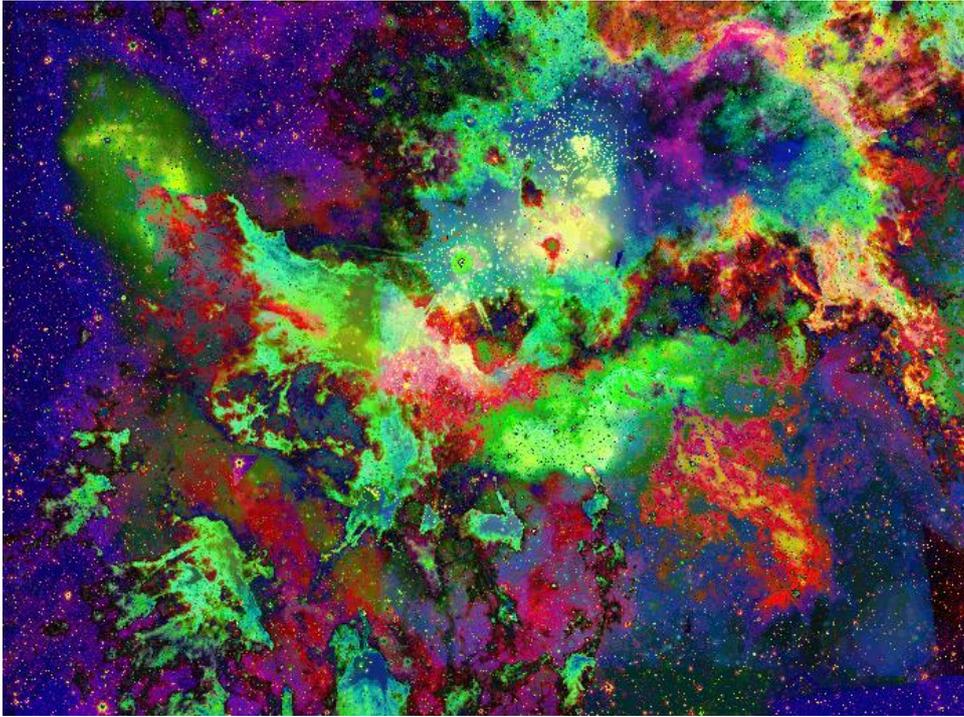


Figure 1. The multispectral electroweak cosmic burst on NGC 3372 binary

3. RESULTS

After my iPoster section of the 240th American Astronomical Society got cancelled due to the interception and interruption of the Chinese cyber espionage, and numerous cyber based harassments by the Chinese cyber operations in my Coursera course, possible connections from the Space Command were received during my July Q&A section with course professor, demanding & commanding my gravity research, which underlies the basis of my assertion that the Space Command was premature, currently running with exceedingly high costs for operation. [48] [49] [2] From my experience in the marriage process and being human trafficked, the militarization of religion by the exogenous semantic network with mandatory use of simplified Chinese is due to the interest in cryptographic network and enslaving territorial population by technical designs, which also underlies the global cyber threats & structural stigma by the increasing use of personal & commercial devices with Simplified Chinese dubbed sources from the CPC command chain. [26]

The antimatter electrolysis behind the NGC 3034 multispectral data processing was achieved with the theoretical hypothesis, which is non-computable but in relation to the physical analysis on instrumentation & satellite-signal-computation paradigm. Temporary Turing completeness was achieved in the series of experiment with the conjecture of infinite irrationals in a topologically invariant globular space such as the earth in relation to time: “whenever the square root of n is irrational, $n^{\frac{1}{n}}$ is irrational”. The mathematical hypothesis in fundamental asymmetry with regard to the fifth cosmic force between the BH and WH that I wanted to present in The Science of Consciousness 2022 is given here with “*” place-holding for electromagnetism according to the energy amplitude either in terms of computer or instrumentation:

$$\left(\frac{i-1}{i}\right)^n * \left(\frac{i+1}{i}\right)^n \vee \left(\frac{i-1}{i}\right)^n * \left(\frac{i+1}{i}\right)^{(n-1)} \text{ with reduced power} \quad (1)$$

The physics aspects regarding BH & WH in terms of nuclear (astro)physics & (astro)chemistry will be revisited later where I can feel safe personally with a peace of mind. The results from NGC 3034 imply that information is preserved one way or another for the non-light chemistry process in multi-wavelength surveys. [44]

4. CONCLUSIONS

The “learn from nature and change nature” doctrine of the Marxist-Communist regime is the epistemological basis not only on the ecological destructions in technological abuses, but also the human rights abuses especially concerning the non-heteronormative gender persons. The gross privacy transgressions in personalized commercial devices are organizationally driven with a dictatorial hypothesis in pancomputationalism. It can be overcome with countermeasures on coercion, and the theoretical methodological framework from experience and research on the scientific & technological challenges took the personal issue as inevitable scientific bias in the real world. It has made a cost and defense effective analysis in available options in a near term from the possibilities in satellite-device interactions in a natural science, specifically astronomical science, based modality. In Chinese natural philosophies there have been two extremity of tendencies: “take arms against a sea of troubles” or “go with the flow”, as in 逆势而为 or 顺应自然. The latter is more equivalent to the principles of natural sciences whereas the technological & militant realities in modern China have been in the former discourse of development as a disguise for the Marxist-Communist doctrines. The methodological analysis primarily serves for the risk management in the former domain.

For warring states to have a human-centred paradigm from the enlightenment tradition in technological designs are the least likely events, especially with modern breaches in semantic security & heterogenous semantic network developments. The exogenous I/O command chain(s) of the communist blocks are making an alternative result-orientated loop by human organizations. The over-concentration on gross privacy breaches is a least effective strategy in risk mitigation, and can further deteriorate psychological and mental health. The development on new scientific & technological innovations is the better way in moving forward with the consistent existence of imminent threats, which justified in part computational sociology for defense loops. The industrial collaborations with the scientific community can have a positive sum effect with the current state of global economy with ongoing economic exploitations by adversary states in the supposedly liberal international institutions. Quantum information’s potential with decentralized networks can render the most sophisticated topologies that is theoretically unhackable even by satellite attacks, with agility. However, the criminal, realpolitik, and power political structural-realist-offensive threats in the liberal institutions is not something that can be solved merely by cyber security.

APPENDICES

Appendix A

The campaign used antiterrorist rhetorics on blocking, misinformation, disinformation, and etc. for its Firewall, and suppressing the domestic population. This was addressed by the Obama administration and Trump administration. Only that it used the codename Poseidon as in maritime aspects in military aggressions not known to the P.R.C. propaganda media environment. This video containing descriptions of the criminal activities was distorted with the narrative of anti-terrorism. [36] The transcription of the evidence is as followed:

Mono: So wonderful the cities and people. But tonight, it will explode accompanying the melodies.

I know those police have been after me and surveillance on me. But I am not afraid.

Because the machines they use have more than ten ports.

To install, they will need more than 30 minutes. And I know how they install.

V.O. To capture an international terrorist, authorize relevant departments using Shanghai Chengxin Technology Ltd. product coded Poseidon, in order to collect important criminal evidence, preventing terrorist attack, and capture the terrorist in the mean time.

Mono: Those police may be sorting through their cables.

“Police”: IP positioning completed. Trojan horse implemented.

“Police” 2: Obtain all files in the hard disk, and all information in the email. Want to capture me with IP?

“Police”: Capturing information disconnected, showing using Internet with VPN.

“Police” 2: Does it raise his suspicions?

“Police”: Rest assured, this newest model can disconnect VPN. I mean, only this one.

Mono: You have to know, I used to be professional. Professionals tell you, how information cannot be stolen. Just as I told you, not downloading .exe files. Apart from preventing Trojan horse, using VPN is not to let the other persons get my IP (address). Right, and your smartphones. Don't root your phones, don't download untrusted sources. Using WeChat? Go take care of your babies. Terrorists use our professional software.

The hotels now are so considerate.

“Police”: Found that our prey is using the energy bank connecting his phone.

“Police” 2: Get all data from his phone. **Mono:** Setting passwords long enough surpassing 20 characters combinations won't get decoded by them. Want to get my plaintext, it takes efforts. Do you know, exploding this city for me is merely a game racing time with the police. So easy.

“Police”: What to do if the target tries VORTEX? Previous tasks cannot prevent VORTEX. We are running out of time.

“Police” 2: This machine called Poseidon is omnipotent. It can be customized.

“Police”: Fetched one email from him.

Mono: Damn the network server. When I think that the whole world is focusing on this wonderful moment, none is a problem.

Hahahaha, mass casualties. Human numbers? Domestic news are too ambiguous. Useless.

(You have a new message, please check.)

Mono: Tell you what is a professional. Professional is ...

“Police”: You’re arrested.

Dialogue: A tea and release? I am used to it. Good luck. I heard SOHO has a lot of casualties.

“Police” 2: Mr., today apart from you having casualties, everywhere else is peaceful.

“Police”: Fetched one email.

“Police” 2: Poseidon is especially useful because it can fake information. Don’t think you are professional; we are more professional.

Appendix B

With all the telecommunication interactions, I only misspoke the Chinese wording on the address to China Merchant Bank staff after the labor trafficking with financial harms and police violence that was possibly talking to the military intelligence operators of the PLA. [11] Albeit I may experience some post-traumatic stress disorder with the ceaseless harassments and some Stockholm syndrome due to a lifetime exposure in such an environment, the fraudulent “psychological” concepts with Chinese ethnic nationalism ideologies and possible broadcasting / telecommunication machine calculations in semantics evidence the criminal behaviours of the adversary coercions & offline based threats in *Figure 2*. The content of the poison pen letter is evidenced in *Figure 3* with transliteration. Since USPS does not operate overseas unless by the U.S. government, and the poison pen letter’s tracking number has no record in the online queries, the envelope can be counterfeited from the Chinese operations in forgery and counterfeiting American government postals for coercion. The use of traditional Chinese in the letter evidences the threats being on power political, realpolitik, and geopolitical aggressions on Taiwan. The contents of the poison pen letter evidences that my caoyang2609@icloud.com email contents were read by the PLA intelligence especially concerning my communications with my husband. [11] One of my Gmail account was breached with harassment emails and the other possibly in covert surveillance. I have been playing sexism with the fake accounts on gay social media with the American officials’ profiles whose privacy had been breached & transgressed, which can be behind the reasons that the paranoid and unscientific term “erotomania” is used in the poison pen letter. The intent of the poison pen letter could have been to try to influence my social interactions from my birth mother whose house currently serve as my safe place, whereas I have a liberal family education from my birth mother who signed the DS60 in front of a Chinese notary regarding my married name in the year 2020. As was evidenced, I have already taken the appropriate step for marriage based on the establishment clause and is recognized in various occasions. [11] The poison pen letter has also threatened the persons I have contact with in isolating me and in isolating the aforementioned persons from corroborating with the truths I have been communicating about - with evidence, investigations, researches, and academic & scientific practices. And with the mass surveillance and gross privacy intrusions, yes, no one can be “strangers” for them. The USCIS envelope dated prior to the poison pen letter has signs of being opened by the Chinese border control personnels evidenced in *Figure 4*, which violates privacy laws and is a common practice in the Communist territory.

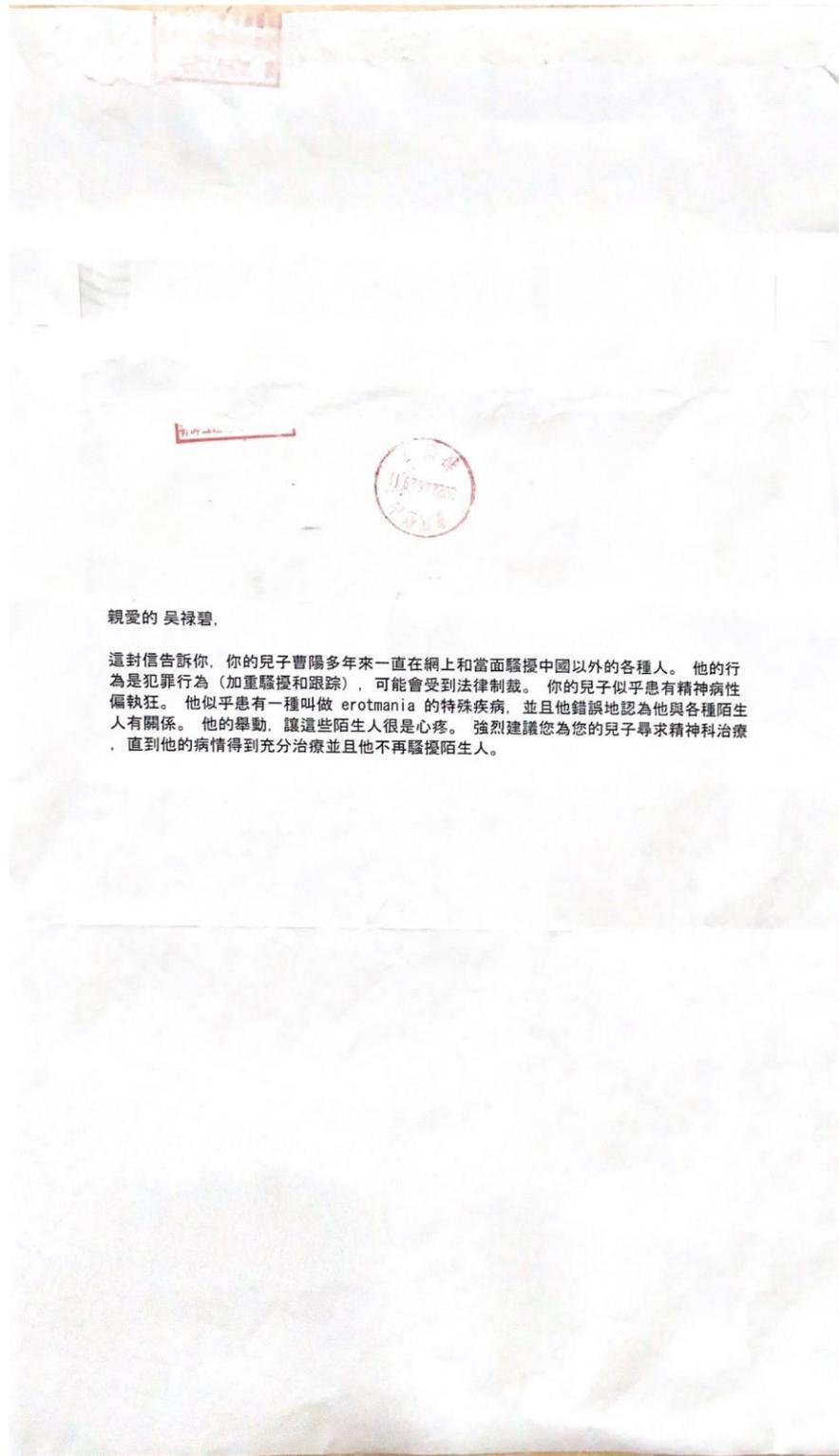


Figure 3. The content of the poison pen letter addressed to my birth mother

Transliteration:

Dear Wu, Lubi,

This letter is to tell (warn) you, your son Cao, Yang for years has been harassing various persons who are not Chinese both online and in person. His behaviours are criminalistic (worsened harassment and stalking), and are possible to be sanctioned by law. Your son seems to have psychological paranoia. He seems to have a disease called erotomania, and he falsely thinks he has relationships with various strangers. His behaviours have hurt the strangers. Intensely recommend you to seek psychotherapy for your son, until his illness is fully treated and never again harassing strangers.



Figure 4. The USCIS envelop with signs of being opened by the P.R.C. border control with COVID-19 rhetorics

Appendix C

I was working on several pieces of manuscripts on Manuscript APP before it turned into I/O. The relevance between interception with the Chinese iCloud backend access and such sudden changes was not determined, and I had to access the original manuscripts with text editor. I noticed the strings of encryption in the coding designs and no other covert operation evidences were detected. The non-ionizing radiation in BH and WH formation in BH physics was what I wanted to tell about in instrumentation safety and the basis on my assertion that Space Command is premature before the completed justifications on a theory of gravity beyond GR. The manuscript was dated on May 2021, and the possible interceptions were the motivation behind the WH observation experiment for the evidence to lay the basis of the judgement. [37]

The fundament of interferometry on instrumentation is on the classical Einstein's equation in solar mass units, and the speed of light either in GR or special relativity. The fundamental concept in reading space telescopes is special relativity with beam power at DCE start, and GR is restrained by gravitational clocking. Gravitational clocking is an important aspect in astronomy, but for BH studies it creates problematic dichotomy in reference systems especially with the improvements on instrumentation with space telescopes. The philosophical part of the mathematical physics approaches with equations is with decimal and mathematical signs. What is meant by a BH's singularity is diverse in disciplines. The chemical part of solar mass is an established unit with Einsteinium GR in astronomy and time is defined with solar clocks. This is plausible for communicational purposes and astronomy within the solar system. A BH has mass and the mass ought to be in galactic units. No matter of the beam power in outer space surveys the speed of light in vacuum is a constant with technology in nature. In other words, the advancement in space telescope instrumentation shortened the distance and planetary limitations posed by solar and lunar eclipses. It increased the efficacy of galactic surveys with wider reference frames. Therefore, the basic notion of BH singularity is its angular momentum based on the location of source beam. The samples used in this research is from Hubble Space Telescope and Chandra Space Telescope (low energy in X-ray). They constitute a geocentric reference system. Eddington limit is thus used in a basic notion that a BH has mass. It is the first derivative from special relativity $L_{accretion} \equiv \epsilon \dot{M} c^2$. The first meaning of BH singularity is the visible accretion luminosity on light particle chemistry. In other words the light source coming from a BH in all wavelengths came from light particle chemistry. With this principle the accretion luminosity of NGC 3034 from 700-6,000 eV was processed. [17] In a geocentric reference system Jonathan Nordebo summarized a linear GR event in a 4-vector metrics in Euclidian space. The Lorentz transformations used in the paper provided a basic solution for elementary particle physics and light particle chemistry. This is to say, in geocentric reference frames solar clocks in solar energy is plausible for perception of a BH distance within the light chemistry limit without the account for light loss on the accretion disk. Since a black body deflects light, the singularity of a BH event horizon signifies current knowledge and deployed application in the solar system. The basic solution composes of an empirical energy momentum binary loop of a BH event horizon with charge. It is important to notice that an event horizon doesn't determine the mass of a black hole. The prerequisite of a black body having mass is solar GR, and the theoretical prerequisite (hypothesis) was implied in the equation. [43]

$$\hat{i}h \frac{\delta}{\delta t} \psi(x, t) = \left[-\frac{\hat{h}^2}{2m} \frac{\delta}{\delta x^2} + V(x, t) \right] \psi(x, t)$$

Appendix D

On July 14, 2022 in Beijing when I tried to have an interview for the pending case and also confirm the integrity of the USCIS mail, I sensed a sudden DoS attack inside of the American Embassy in Beijing during the interview, and several persons on WeChat reported to me on the malfunctioning of their electronic devices. *Figure 5* is an example of the polymer blast apart from another who reported on broken phone screen. The dictatorial reliance on artificial intelligence and the big data technologies can be a contributing factor for the applications of electronic warfare. [11] The PLA's conduct of Great Firewall & electronic warfare pose non-traditional nuclear threat to Beijing itself and imposes transgressions to the civil society's right to health. It has been causing public health crisis with such operations and grossly violating the Geneva Conventions. Detailed accounts of the investigations & research will be given elsewhere.



Figure 5. The electronic warfare's shockwave impact on the civil society in Beijing

ACKNOWLEDGEMENTS

I would like to thank the Biden administration for building back better with timely responses on the human rights issues that constitute the building-blocks of American democracy. I appreciate the reviewers' feedback on the first manuscript written in a short amount of time, and the

editorial office's feedback. I owe my gratitude to the black hole studies & quantum physics community, and the online astronomy course instructors for their understanding in my attendance of the courses accompanied by intrusive signals that sometimes impacted the normal runnings of the *Coursera.org* community.

REFERENCES

- [1] Cox, Robert W. (1997) *The New Realism: Perspectives on Multilateralism and World Order*, Multilateralism and the UN System, The United Nations University Press. <https://doi.org/10.1007/978-1-349-25303-6>.
- [2] Pachankis, Yang (2020) "Lateralism – The Globalization of US Hegemony after World War II ", *Dissertation*, Zenodo. <https://doi.org/10.5281/zenodo.6428349>.
- [3] Rots, Arnold H. *et al.* (2015) "Representations of time coordinates in FITS. Time and relative dimension in space", *Astronomy and Astrophysics*, Vol. 574, pp. A36. <https://dx.doi.org/10.1051/0004-6361/201424653>.
- [4] Greisen, E. W. & Calabretta, M. R. (2002) "Representations of world coordinates in FITS", *Astronomy and Astrophysics*, Vol. 395, pp. 1061-1075. <https://dx.doi.org/10.1051/0004-6361:20021326>.
- [5] Calabretta, M. R. & Greisen, E. W. (2002) "Representations of celestial coordinates in FITS", *Astronomy and Astrophysics*, Vol. 574, No. 3, pp. 1077 - 1122. <https://doi.org/10.1051/0004-6361:20021327>.
- [6] Greisen, E. W., Calabretta, M. R., Valdes, F. G., and Allen, S. L. (2006) "Representations of spectral coordinates in FITS", *Astronomy and Astrophysics*, Vol. 446, No. 2, pp. 747 - 771. <https://doi.org/10.1051/0004-6361:20053818>.
- [7] Kaldor, Mary (2011) "War and Economic Crisis", in *The Deepening Crisis: Governance Challenges after Neoliberalism* (ed Calhoun, Craig and Derluquian, Georgi), New York University Press, New York and London, pp 109-133. New York University Press. <https://doi.org/10.18574/nyu/9780814772805.003.0006>.
- [8] Yan, Guo (2005) "National Identification in the Context of Globalization", Doctoral Thesis, International Strategy Institute, Party School of the Central Committee of CPC.
- [9] Pachankis, Yang (2020) "A Multicultural Retrospective on Endogenous Chinese Sino-Centric Civilizational Becoming", *Doctoral Thesis*, Zenodo. <https://doi.org/10.5281/zenodo.6847355>.
- [10] Yuan, Jing-Dong (2007) "Culture matters: Chinese approaches to arms control and disarmament", *Contemporary Security Policy*, Vol. 19, Iss. 1, pp. 85-128. <https://doi.org/10.1080/13523269808404180>.
- [11] Pachankis, Yang (2022) "Epistemological Extrapolation and Individually Targetable Mass Surveillance: The Issues of Democratic Formation and Knowledge Production by Dictatorial Controls", *International Journal of Innovative Science and Research Technology*, Vol. 7, Iss. 4, pp. 72–84. <https://doi.org/10.5281/zenodo.6464858>.
- [12] Kahneman, Daniel (2011) *Thinking, Fast and Slow*, Farrar, Straus and Giroux. ISBN: 978-0374275631.
- [13] Pachankis, John E. (2007) "The psychological implications of concealing a stigma: a cognitive-affective-behavioral model", *Psychological bulletin*, Vol. 133, Iss. 2, pp. 328–345. <https://doi.org/10.1037/0033-2909.133.2.328>.
- [14] Bollas, Christopher (1993) "The Fascist State of Mind", in *Being a Character: Psychoanalysis and Self Experience*, pp. 193-217.
- [15] Lindquist, Kristen A., MacCormack, Jennifer K., and Shablack, Holly (2015) "The role of language in emotion: predictions from psychological constructionism", *Frontiers in Psychology*, Vol. 6:444. <https://doi.org/10.3389/fpsyg.2015.00444>.
- [16] Pachankis, Yang (2022) "Reading the Cold War through Outer Space: The Past and Future of Outer Space", *International Journal of Scientific & Engineering Research*, Vol. 13, Iss. 6, pp. 826-829. <https://doi.org/10.14299/ijser.2022.06.03>.
- [17] Pachankis, Yang (2021) "Research on the Kerr-Newman Black Hole in M82 Confirms Black Hole and White Hole Juxtapose (Thermonuclear Binding)", *Academia Letters*, Article 3199. <https://doi.org/10.20935/AL3199>.

- [18] Pachankis, Yang Immanuel (2022) "White Hole Observation: An Experimental Result," *International Journal of Innovative Science and Research Technology*, Vol. 7, Iss. 2, pp. 779-790. <https://doi.org/10.5281/zenodo.6360849>.
- [19] Ou, Yang (2012) "High-speed PCIe SSD Studies & Realization Based on RAM," dissertation, National University of Defense Technology affiliated with the PLA.
- [20] Zhang, Wei & Jiang, Zihao & Chen, Zhiguang & Xiao, Nong & Ou, Yang (2021) "NUMA-Aware DGEMM Based on 64-Bit ARMv8 Multicore Processors Architecture," *Electronics*. 10. 1984. <https://doi.org/10.3390/electronics10161984>.
- [21] Gill, Stephen (1997) *Globalization, Democratization, and Multilateralism*, Multilateralism and the UN System, Palgrave Macmillan. <https://doi.org/10.1007/978-1-349-25555-9>.
- [22] Biddle, Tami Davis (2020) "Coercion Theory: A Basic Introduction for Practitioners", *Texas National Security Review*, Vol. 3, Iss. 2. <http://dx.doi.org/10.26153/tsw/8864>.
- [23] Nightingale, Edmund B., Hodson, Orion, McIlroy, Ross, Hawblitzel, Chris, and Hunt, Galen (2009) "Helios: heterogeneous multiprocessing with satellite kernels", *Association for Computing Machinery*, Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles, pp. 221–234. <https://doi.org/10.1145/1629575.1629597>.
- [24] Chen, Bo & Xiao, Nong & Liu, Fang & Ou, Yang & He, Wanhui (2015) "An SSD Architecture Optimization Method for RAID Parallels", *Computer Engineering and Science*, 7(2014). <https://doi.org/10.3969/j.issn.1007-130X.2014.07.003>.
- [25] Wu, Lizhou & Xiao, Nong & Liu, Fang & Du, Yimo & Li, Shuo & Ou, Yang. (2015) "Dysource: A High Performance and Scalable NAND Flash Controller Architecture Based on Source Synchronous Interface", The 12th ACM International Conference, 1-8. <https://doi.org/10.1145/2742854.2742873>.
- [26] Hatzenbuehler, Mark L. (2016) "Structural stigma: Research evidence and implications for psychological science," *American Psychologist*, 71(8), pp. 742–751. <https://doi.org/10.1037/amp0000068>.
- [27] Wulf, Norman A. (1985) "Outer Space Arms Control: Existing Regime and Future Prospects", *Documents on Outer Space Law*. 9. <https://digitalcommons.unl.edu/spacelawdocs/9>.
- [28] Alhomidi, Mohammed & Reed, Martin (2014) "Attack Graph-Based Risk Assessment and Optimisation Approach", *International Journal of Network Security & Its Applications (IJNSA)*, Vol.6, No.3, pp. 31-43. <https://doi.org/10.5121/ijnsa.2014.6303>.
- [29] Ching, Ke Wan & Singh, Manmeet Mahinderjit (2016) "Wearable Technology Devices Security and Privacy Vulnerability Analysis", *International Journal of Network Security & Its Applications (IJNSA)*, Vol.8, No.3, pp. 19-30. <https://doi.org/10.5121/ijnsa.2016.8302>.
- [30] Pachankis, Yang I. (2022) "Physical Signals and their Thermonuclear Astrochemical Potentials: A Review on Outer Space Technologies," *International Journal of Innovative Science and Research Technology*, Vol. 7, Iss. 5, pp. 669-674. <https://doi.org/10.5281/zenodo.6618334>.
- [31] Pachankis, Yang (2007) "'Ménage à trois' -- A Legal Discourse on Marriage Inequality," *Academia*, theatrical work originally debuted in Communication University of China. <https://www.academia.edu/video/joxGK1>.
- [32] Pachankis, John E. *et al.* (2021) "Welcome to Chechnya Film Screening and Panel Discussion," *Yale School of Medicine*. <https://medicine.yale.edu/media-player/welcome-to-chechnya-film-screening-and-panel-discussion/>.
- [33] Pachankis, Yang Immanuel (2022) "The Experimental Psychology on Cognitive-Affective-Behavioral Process," *Zenodo*. <https://doi.org/10.5281/zenodo.6844274>.
- [34] Pachankis, John E. *et al.* (2021) "Brief online interventions for LGBTQ young adult mental and behavioral health: A randomized controlled trial in a high-stigma, low-resource context," *J Consult Clin Psychol*. 2020 May; **88**(5): 429–444. <https://doi.org/10.1037%2Fccp0000497>.
- [35] Pachankis, Yang (2022) "The Militarization of Cyber Domains by the People's Liberation Army and P.R.C. Party-control of Governmental Apparatus," *Zenodo*. <https://doi.org/10.5281/zenodo.6837265>.
- [36] Pachankis, Yang (2022) "Secondary evidence for totalitarian technological abuses," *Academia*. <https://www.academia.edu/video/kOv0Jl>.
- [37] Pachankis, Yang (2022) "White Hole on the Trifid Nebula," *Zenodo*. <https://doi.org/10.5281/zenodo.6426887>.
- [38] Oberg, K. (2021) "Origins Of Astrochemical Complexity," *Bulletin of the AAS*, 53(6). <https://baas.aas.org/pub/2021n6i200p01>.
- [39] Menzies, Gavin(2003) "1421, The Year China Discovered America," *Perennial*, ISBN: 0-06-054094-X.

- [40] Dvali, Gia & Kuhnel, Florian & Zantedeschi, Michael (2021) "Vortexes in Black Holes," *arXiv*. <https://doi.org/10.48550/arXiv.2112.08354>.
- [41] Run, Dong & Liu, Fang & Xiao, Nong & Chen, Xiang & Ou, Yang (2015) "PCIe SSD I/O Bus Design and Prototype System Research," NCIS 2015, China Computer Federation, pp. 124-130.
- [42] Pachankis, Yang (2022) "The Modern Origins & Sources of China's Techtransfer," *International Journal of Scientific & Engineering Research*, Vol. 13, Iss. 7, pp. 18-25. <https://doi.org/10.14299/ijser.2022.07.01>.
- [43] Pachankis, Yang I. (2022) "Some Concepts of Space, Time, and Lengths in Simplified Chinese* An Analytical Linguistics Approach," *International Journal of Innovative Science and Research Technology*, Vol. 7, Iss. 6, pp. 550-562. <https://doi.org/10.5281/zenodo.6796083>.
- [44] Pachankis, Yang I. (2022) "A Multi-wavelength Data Analysis with Multi-mission Space Telescopes," *International Journal of Innovative Science and Research Technology*, Vol. 7, Iss. 1, pp. 701-708. <https://doi.org/10.5281/zenodo.6044904>.
- [45] Pope, Robert (2021) "Black-light's ability to decode white-light's reality to reveal the complete structure of the evolving universe is the most important discovery in the history of science," *Academia*. https://www.academia.edu/50786003/Black_lights_ability_to_decode_white_lights_reality_to_reveal_the_complete_structure_of_the_evolution_universe_is_the_most_important_discovery_in_the_history_of_science.
- [46] Labanji, Fagbote Olawumi (2022) "Performance Evaluation of Strong Techniques," *Global Scientific Journals*, Vol. 10, Iss. 1.
- [47] Dwyer, Joseph R. (2012) "The relativistic feedback discharge model of terrestrial gamma ray flashes," *Journal of Geophysical Research Space Physics*, Vol. 117, Iss. A2. <https://doi.org/10.1029/2011JA017160>.
- [48] Astronomy: State of the Art (2022) "June 22th, 2022 Live Astronomy Q&A Session with Prof. Chris Impey," Youtube. <https://youtube.com/watch?v=X-A5MAqy7k8>.
- [49] Hartmann, Margaret (2022) "Let's Get to Know Space Force, Trump's Most Misunderstood Creation," *Intelligencer*, Jul. 2022. <https://nymag.com/intelligencer/article/space-force-guide.html>.

AUTHOR

Yang Pachankis

(birth name Yang Cao)

Graduated from Verakin High School of Chongqing with studies in the nuclear & astronomical sciences; He holds a B.A. in Directing (Editing Art & Technology) from Communication University of China where he also attended graduate studies.

He is granted Masters & PhD in Global Governance, Cosmology on *ex post facto* basis.



INTEGRATING ETHICAL, LEGAL AND SOCIAL ASPECTS INTO COMMON PROCEDURE MODELS

Sascha Alpers

FZI Forschungszentrum Informatik, Karlsruhe, Germany

ABSTRACT

Many different procedure models can be applied to the management of software development projects. Such models also consider the ascertainment and management of requirements – based on very different agile or classic approaches. The framework provided in particular by ethical aspects, legal constraints and social technology design issues (ELSA or ELSI) is not explicitly addressed in procedure models, which is why approaches such as the IEEE Standard Model Process for Addressing Ethical Concerns during System Design (IEEE7000-2021) have been developed. However, the lack of explicit integration of these issues into common process models such as SCRUM or V-ModellXT implies a lack of necessary space for reflection on ELSA within development projects. The article discusses this problem and highlights possible solutions for further discourse.

KEYWORDS

Procedure model, ethics, law, social technology design.

1. INTRODUCTION

Many different procedure models can be applied to the management of software development projects. Such models also consider the ascertainment and management of requirements – based on very different agile or classic approaches. The framework provided in particular by ethical aspects, legal constraints and questions of social technology design (ELSA, or ELSI) is not explicitly addressed in the procedure models. Since software is developed in part for a large number of future use cases – some of which have little specific context – this makes the challenge faced by procedure models more complex (for example, those cases affected are not yet concretely known and therefore cannot be involved; instead, other ways of taking their interests into account must first be identified). The lack of explicit integration of ELSA considerations into common procedure models such as SCRUM or V-Modell XT implies a lack of necessary space for reflection on ELSA within development projects. This contradicts the importance of ELSA aspects as found in the ethical guidelines of the German Informatics Society [1] and other scientific [2] and social [3] sources. This article discusses this problem and highlights possible solutions for further discourse, including in a workshop format.

The second section presents the current state of the art in science and technology. The basics of process models are summarised, after which the section looks at how ELSA is currently considered in existing general procedure models. Procedure models that specialise in ELSA issues are also considered.

The third section presents four theses on the future consideration of ELSA in software development projects. These theses are intended to stimulate further discussion and lead to the further development of a systematic consideration of ELSA.

2. STATE OF THE ART IN SCIENCE AND TECHNOLOGY

2.1. Procedure Models

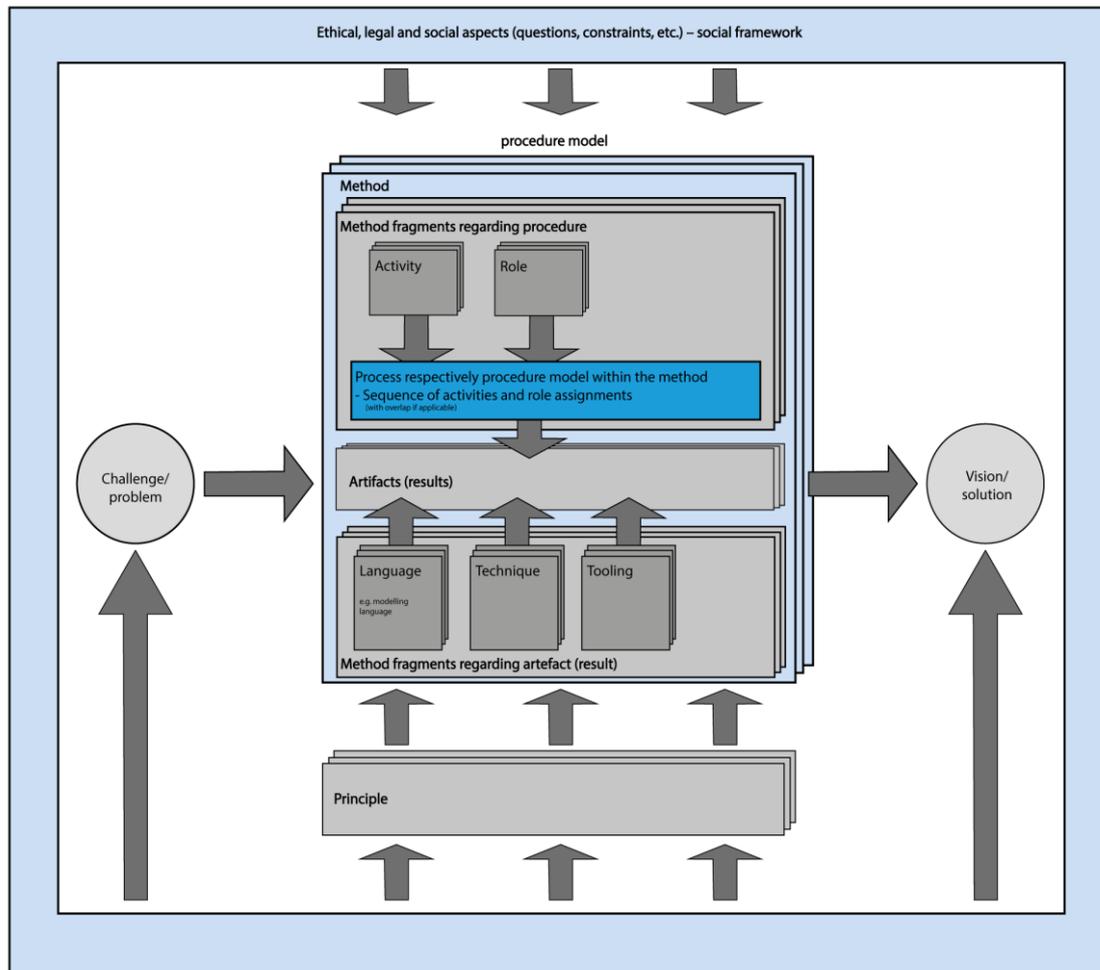


Figure 1. Ethical, legal and social aspects as a framework for procedure models (based on [4])

Software development projects are usually structured through organisational and, if necessary, project-specific adaptations of existing process models. As Figure 1 shows, procedure models are used – based on a certain starting point – to design the path to a particular goal. According to [5], a procedure model (also known as a ‘procedure strategy’) combines different methods or method fragments. The following components define a method [cf. 6]: The process (sometimes also known as a ‘procedure model within the method’) defines which activities are conducted in which order (with temporal overlap where applicable) and under which conditions by which roles. Activities and roles are shown separately in the figure; the additional boxes in the background of the figure show that there are several activities or roles within a single method. Modelling experts or moderators typically conduct these activities, with one or more roles able to be involved in performing an activity.

Artefacts (results) are generated by activities or within the framework of activities. In turn, some artefacts are also required as input in order to be able to perform other activities [6]. The following components are relevant here:

- **Language:** all results must be described in a certain language. This may take the form of natural language or a specific, potentially specialised format. Language definitions range from sentence templates [7, p. 57ff.] to tables and diagrams and even formalised models. The language should not only be syntactically defined for a highly formalised application, but its semantics should be clearly specified [5].
- **Technique:** this refers to ‘the respective regulation for creating (and thus documenting) the results’ [6, p. 88, own translation].
- **Tools:** these can be used within a method to, for example, support the technique. Tools may support different activities as part of a method, for example the tool-supported moderation of action planning (e.g. using Miro Board), enterprise modelling (e.g. using Horus Business Modeler), process modelling (e.g. using Camunda Modeler) and formulating user stories.

For software development projects, there are now a number of process models which contain these components to varying degrees and levels of intensity. A fundamental distinction should be made between the different philosophies of agile and classic, of which different representatives are used in practice [8].

2.2. Ethical, Legal and Social Aspects (ELSA) in Process Models

Procedure models for software development projects have no structural specific anchoring in terms of ethical, legal and social aspects. Anchoring through explicit elements (e.g. specific activities, roles or artefacts) does not exist in typical procedure models. Instead, ethical, legal and social aspects are typically considered when ascertaining requirements (if at all) and when ascertaining, agreeing on and documenting (concrete) non-functional requirements. The overarching consideration within the project – especially in terms of follow-up – is then factored into the requirements management process along with other non-functional requirements. This approach is used for other specific aspects (such as IT security) as well.

Due to the fact that functional requirements only gradually emerge during the project, agile process models call for an individual (functional) requirement to be ethically, legally and socially coordinated as an additional quality requirement. Within the Scrum process model, the quality requirements are defined under the ‘Definition of Ready’ [9]. This ‘Definition of Ready’ may then also contain requirements regarding the coordination of a (functional) requirement for ELSA. This approach within agile projects is also used for other specific aspects (such as IT security).

For agile process models, the agile manifesto [10] can be seen as a summary of the core philosophy surrounding the process. It was formulated in 2001 by 17 signatories as the lowest common denominator of various agile process models. The basic ideas are summarised in the form of four values and twelve principles. The first and third values are formulated as follows: The manifesto signatories value ‘individuals and interactions more than processes and tools’ and ‘collaboration with the customer more than contract negotiation’ [10]. The orientation towards natural persons promotes a human-centred approach as it also exists in the field of social technology design [cf. 11]. The principles of the agile manifesto turn the focus onto customers and subject matter experts (often business users) as well as developers and go into this in more detail. Other stakeholders in the field of technology design are not explicitly mentioned.

The option to consider ELSA-related aspects in a structured manner are indicated by special standards. However, a procedure model based on a special standard cannot be integrated directly into a typical software engineering procedure model. Instead, this model was created as an independent, autonomous procedure and can be used in projects if those responsible are aware of this and wish to implement it. An example of such a special standard is the IEEE standard ‘IEEE Model Process for Addressing Ethical Concerns during System Design’ (IEEE 7000-2021), which was first published in 2021. IEEE 7000-2021 provides for two phases: In the first phase – concept exploration – the concept of use and the context are explored in order to determine and prioritise ethical values. The second phase – definition of ethical requirements – begins with concept research and continues into the development stage. A design process reflecting ethical considerations is also part of this stage.

In the context of business ethics, there are various approaches for companies or their managers to arrive at decisions that take ethics into consideration. However, these approaches usually refer to the business model level rather than the level of technology design. Such approaches do provide the opportunity to learn about ethical considerations surrounding (information) technology [12] in general and, more specifically, the integration of ELSA into software engineering procedure models.

3. THESES FOR FURTHER DEVELOPMENT

The previous section described the current state of science and technology. The question is how science and technology will continue to evolve and how they can be actively developed. The following theses are intended to contribute to the discussion, to serve as a catalyst for the work performed by expert groups on procedure models and project management, and to inform science and practical considerations in general.

- Thesis 1: For the effective and efficient inclusion of ethical, legal and social aspects, it is not sufficient to consider them from a general perspective as a (social) framework for procedure models or development projects. This is because it fails to incorporate ELSA as a ‘standard’ consideration and does not sufficiently support either those responsible for the project or those actually carrying it out.
- Thesis 2: ELSA is too different (e.g. compared to other requirements) to be considered purely as ‘incidental’ in process models (e.g. requirements engineering) not specifically designed for this purpose (whether socially relevant, partly complex, etc.).
- Thesis 3: Further structural anchoring of ELSA is required in standard procedure models: specific activities (e.g. quality gates with ethics checks, involvement of ‘affected parties’ such as employee representatives, etc.), specific roles (e.g. ethics officers), concrete anchoring of activities and roles within a procedure model, specific artefacts (e.g. value register), and so on. Depending on the procedure model, this ensures that ELSA receives the requisite attention across the entire life cycle of systems. This also takes into consideration the fact that ELSA-related matters or requirements may change during the life cycle and have to be implemented, e.g. for ‘maintenance’ purposes.
- Thesis 4: Activities are required that are independent of any aspect of a procedure model. This includes the formation and maintenance of an organisation’s core values. These values can then be used in the organisation’s projects and serve as a working basis for coordination as part of cross-organisational projects (this will be based around non-negotiable values where compliance is mandatory if a project partnership with other organisations is to be established). Other measures include making certain professional groups in general and employees aware of ELSA through suitable education and training efforts (cf. GI Ethics Standard). In general, the implementation of thesis 4 will lead to a broad cultural change.

4. CONCLUSIONS

Those responsible for procedure model projects and those tasked with taking action at various levels continue to be called on to accept the requisite responsibility for ELSA or to craft design proposals for ethically responsible, legally permissible and socially good procedures and solutions. The topics presented in this publication are intended to prompt discussion and work regarding the systematic consideration and integration of ELSA into software projects. Discourse on this matter can be informed by designers of procedure models as well as prototypical but well-considered adaptations of procedures in specific contexts (company, project, etc.). Such discourse can be bolstered by scientific findings as desired, with general lessons able to be extracted from the respective contexts.

A broad exchange of integration options is required for the sustainable, cross-contextual integration of ELSA into procedure models in software projects. This can be fostered by seeking out individual contributions such as project reports on adapted procedures. It is hoped that further discourse will lead to adapted, context-independent process models. A first step towards adapting established process models in practice could then be to conduct training courses on the appropriate integration of ELSA into existing process models.

ACKNOWLEDGEMENTS

This publication was developed as part of the project ‘Competence Centre KARL – Artificial Intelligence for Work and Learning in the Karlsruhe Region’. This research and development project is funded by the German Federal Ministry of Education and Research (BMBWF) within the program ‘The Future of Value Creation – Research on Production, Services and Work’ (funding number 02L19C250) and is managed by the Project Management Agency Karlsruhe (PTKA). The author is solely responsible for the content of this publication.

REFERENCES

- [1] Gesellschaft für Informatik e.V. (GI) (2018): ‘Ethical Guidelines of the German Informatics Society’, <https://gi.de/ethicalguidelines>, last accessed 2022-06-14
- [2] Gotterbarn, D.; Miller, K.; Rogerson, S. (1997): ‘Software engineering code of ethics’. *Communications of the ACM*, 40/11, New York, Association for Computing Machinery, pp. 110-118, <https://doi.org/10.1145/265684.265699>.
- [3] Himmer, N. (2019): ‘Computer und die Moral – Philosophische Nachhilfe für Nerds’, Frankfurter Allgemeine Zeitung Online, <https://www.faz.net/-gyl-9ibi7>, last accessed 2022-06-14
- [4] Alpers, S. & Karle, T. & Schreiber, C. & Schönthaler, F. & Oberweis, A. (2021): ‘Process Mining bei hybriden Vorgehensmodellen zur Umsetzung von Unternehmenssoftware’, *Informatik Spektrum*, Vol. 44, pp. 178–189. <https://doi.org/10.1007/s00287-021-01359-7>
- [5] Brinkkemper, S. (1996): ‘Method engineering: engineering of information systems development methods and tools’. *Information and Software Technology*, 38/4, pp. 275–28, [https://doi.org/10.1016/0950-5849\(95\)01059-9](https://doi.org/10.1016/0950-5849(95)01059-9)
- [6] Winter, R. (2003): ‘Modelle, Techniken und Werkzeuge im Business Engineering’. *Business Engineering*. Springer, Berlin, Heidelberg, pp. 87–118. https://doi.org/10.1007/978-3-642-19003-2_5
- [7] Pohl, K.; Rupp, C. (2015): ‘Basiswissen Requirements Engineering’ (4. eds). dpunkt.verlag GmbH.
- [8] Kuhrmann, M.; Linssen, O. (2014): ‘Welche Vorgehensmodelle nutzt Deutschland?’ *Projektmanagement und Vorgehensmodelle 2014*. Bonn: Gesellschaft für Informatik e.V., pp. 17-32.
- [9] Dalton, J.: Definition of Ready. *Great Big Agile*. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4206-3_26
- [10] Beck, K. & Beedle, M. & Bennekum, A. & Cockburn, A. & Cunningham, W. & Fowler, M. & Grenning, J. & Highsmith, J. & Hunt, A. & Jeffries, R. & Kern, J. & Marick, B. & Martin, R. &

- Mellor, S. & Schwaber, K. & Sutherland, J. & Thomas, D. (2001): Manifesto for Agile Software Development, <https://agilemanifesto.org>
- [11] Klose, E. & Ni, I & Schmidt, L. (2019): ‘Mit benutzerzentrierter Entwicklung zur Integration von sozialen Aspekten in die Projektarbeit’. In *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft* (Workshop-Beiträge). pp 447-463. https://dx.doi.org/10.18420/inf2019_ws49
- [12] Ulrich, P. (1998): ‘Integrative Wirtschaftsethik — eine Heuristik auch für die Technikethik?’ In: Lenk, H. (ed.); Maring, M. (ed.): *Technikethik und Wirtschaftsethik*. VS Verlag für Sozialwissenschaften, Wiesbaden. https://doi.org/10.1007/978-3-322-97402-0_3

AUTHORS

Sascha Alpers studied Information Engineering and Management at the Karlsruhe Institute of Technology (KIT), specialising in information and knowledge systems, information services in networks, and business processes and organisation. In 2019, he was awarded his doctorate by the KIT Faculty of Economics. At the “FZI Forschungszentrum Informatik” (FZI Research Centre for Information Technology, www.fzi.de), Sascha Alpers researches issues of digital sovereignty surrounding business processes and information systems in his role as a department manager in the software engineering research division. He also heads the FZI Living Lab Software Innovations. Current FZI projects include SDIKA (manager of FZI sub-projects, www.sdika.de) and KARL (responsible for the ethical and legal issues work package, www.kompetenzzentrum-karl.de).



VOICE CHATBOT FOR HOSPITALITY

Sagina Athikkal and John Jenq

Department of Computer Science, Montclair State University, NJ USA

ABSTRACT

Chatbot is a machine with the ability to answer automatically through a conversational interface. A chatbot is considered as one of the most exceptional and promising expressions of human computer interaction. Voice-based chatbots or artificial intelligence (AI) devices transform human-computer bidirectional interactions that allow users to navigate an interactive voice response (IVR) system with their voice generally using natural language. In this paper, we focus on voice based chatbots for mediating interactions between hotels and guests from both the hospitality technology providers' and guests' perspectives. We developed a hotel web application with the capability to receive a voice input. The application was developed with Speech recognition and deep synthesis API for voice to text and text to voice conversion, a closed domain question answering (cdQA) NLP solution was used for query the answer.

KEYWORDS

Natural Language Processing, Chatbot, Voice Based Digital Assistants, Closed Domain Question Answering.

1. INTRODUCTION

A chatbot is a programming interface that simulates the conversation or "chatter" of a human being through text or voice interactions. Nowadays, chatbots are available in almost many aspects of technology, such as mobile assistants, customer services, e-commerce, and smart devices. It is a type of software which can help the users by automating their conversations and interact with the customers through the messaging platforms. These chatbot-virtual assistants are found useful to handle simple, look-up tasks in business-to-consumer and business-to-business environments. Chatbot virtual assistants are helpful not only to make use of support staff time but also beneficial in providing a level of customer service when the supporting agents aren't available [1]. Chatbots interpret and process user's words or phrases giving them an instant pre-set answer [2]. The most important aspect of implementing a chatbot is selecting the right natural language processing (NLP) engine. If the user interacts with the bot through voice, for example, then the chatbot requires a speech recognition engine. Similar to regular apps, chatbots also have an application layer, a database, APIs, and Conversational User Interface (CUI)[2]. There are structured and unstructured conversations. Chatbots built for structured conversations are highly scripted, it simplifies programming but restricts the kinds of things that the users can ask. In most B2B environments, chatbots are commonly scripted and used to respond to frequently asked questions or perform simple, repetitive calls to action. In sales, a chatbot may be a quick way for sales reps to get phone numbers. For service departments, it assisting service agents in answering repetitive requests. Generally, once a conversation gets too complex for a chatbot, the call or text window will be transferred to a human service agent.

Chatbots such as ELIZA and PARRY were early attempts at creating programs that could at least temporarily fool a real human being into thinking they were having a conversation with another

person. PARRY's effectiveness was benchmarked in the early 1970s using a version of a Turing test; testers only made the correct identification of a human versus a chatbot at a level consistent with making a random guess.

Chatbots have come a long way since then. They are built on artificial intelligence (AI) technologies, including deep learning, natural language processing and machine learning (ML) algorithms, and require massive amounts of data. The more an end user interacts with the bot, the better voice recognition becomes at predicting an appropriate response. We can roughly classify chatbots into three categories: (a) Rule-based, this is the simplest type of chatbots. They require user to make a few selections, such as using drop downs or buttons, to give relevant answers. They are slow but is easy to implement. When many conditions or factors are involved in the knowledge base, this approach may not be the best solution [3]. (b) Intellectually independent chatbots: These chatbots learn from the user's inputs and requests by using Machine Learning. This kind of bots are trained in such a way to understand specific keywords and phrases that triggers bot's reply. They train themselves to understand more and more questions with practice and experience [3]. They spot keywords or phrases and provide predefined answer based on these spotted keywords or phrases. (c). AI-powered chatbots: It combines the best from the rule-based and intellectually independent chatbots. These bots understand free language and make sure they solve the user's problems with a predefined flow. They can switch the conversational scenario when needed and address random user requests at any moment. These chatbots use machine learning, AI, and Natural Language Processing (NLP) to understand and analyse human speech, find the right response and reply in understandable way in a human language.

Overall speaking, chatbots are considered as one of the most advanced and promising aspect of interaction between humans and machines. Chatbot applications helps in smoothening the interactions between the customers and the services [4]. They can enhance and engage customer interactions with less human intervention.

In [5], Dimitrios Buhalis and Iuliia Moldavska explained the importance of voice assistants in the hotel industry. They clearly mention the advantages of voice assistants in hotels outweigh the disadvantages for both hotels and guests. Their findings illustrate that voice-based human-computer interactions bring a range of benefits and voice assistants will be widely deployed in the future. Technology integrations are often complex and costly to set up but it provides significant benefits especially in hotel and tourism industry. As reported by them, guests appreciate the prospective benefits but are concerned with privacy and usability, although tech-savvy consumers are less concerned about privacy when using voice assistants. The findings indicated the direction for the future development of voice technology in hospitality towards multilingualism and modulated offers which can ultimately ensure the overall wider reach of the technology in the hotel industry.

Li, Bai et.al. [6] reviewed a real-world conversational AI and NLP system for hotel booking. Their architecture design includes a frame-based dialogue management system that calls machine learning models for classification, named entity recognition, and information retrieval subtasks. Their chatbot has been deployed on a commercial scale, handling tens of thousands of hotels searches every day. They have also explained various machine learning models that they used for deployment and explained developing an e-commerce chatbot in the travel industry. Adam et.al. [7] describes the significance of chatbots in various fields. They explain how the use of ADCs (Identity, small talk, and empathy) as a common compliance technique, affect user compliance with a request for service feedback in a chatbot interaction. They have examined a randomized online experiment how verbal anthropomorphic design cues and the foot-in-the-door technique affect user request compliance. Their results show that both anthropomorphism as well as the need to stay consistent significantly increase the likelihood that users comply with a chatbot's

request for service feedback. They have commented that social presence mediates the effect of anthropomorphic design cues on user compliance. In [8], Hasanet.al. examine touristchatbot usage intentions in service encounters within the context of a future international travel, assuming continued social distancing. Their results show that automation, habit, social presence, and health consciousness all contributed positively to chatbot usage intentions. Some variations were observed as a function of experiencing government-imposed lockdown. The role of social presence and human qualities in chatbots was weakened when controlling for lockdown and during the trip experience.

2. PROPOSED SYSTEM

Our proposed system recognizes speech on chatbot which uses NLP for interaction on any closed domain system. It uses latest technology scope forward and developed the application in internet. Users can use voice to interact with the web application instead of searching and navigate the website. Any questions related to the hospitality or the hotel web application (like how to use, or where can I find it), or questions related to the business domain can get answered by the chatbot.

The system is trained to answer any question related to hospitality domain. The system is trained with any data set which has information related to the hotel website. It is implemented in a way so it can be easily retrained with any data set.

The proposed system requires following modules: a hotel web site which can host the voice chat bot. Web applications need to capture voice input and get the voice to convert into text. There are various solutions for this. Following are two most used voice to text conversion for web application.

2.1. Speech to Text Engine

Deep Speech is an open-source voice recognition and speech to text engine, which provide a trained model using Baidu's Deep Speech research paper and the model implementation is under Mozilla Public license. The underlying implementation is using Google's Tensor Flow. It come up with two models the acoustic model and the language model. Acoustic model is an end-to-end deep leaning system, and the language model is used to increase the accuracy of the transcription output which is included as separate model. The language model can be customized based on our domain.

For the implementation, we must download the model first. The .pbmm is the acoustic model which is trained based on American English, in behind the scene it uses tensor flow. Scorer is the language model, which is useful for improving the accuracy of the predicted output. For example, using this, it will find out which word is grammatically right in a particular context.

The architecture of the engine was originally based up on Deep Speech: Scaling up end-to-end speech recognition. Currently it is different in many aspects and made it based on recurrent neural network (RNN) which is trained to ingest speech spectrogram and generate English text transcription [9]. Deep Speech model use hybrid model for parallel optimization. Hybrid parallel optimization combines the benefit of asynchronous and synchronous optimization. It allows to use multiple GPUs but doesn't have a problem of incorrect gradient present in asynchronous optimization.

In hybrid parallel optimization initially, it places the model in CPU memory. Then, as in asynchronous optimization, each of the G GPUs obtains a mini batch of data along with the

current model parameters. Using the mini batch each of the GPUs then computes the gradients for all model parameters and sends these gradients back to the CPU. Now, in contrast to asynchronous optimization, the CPU waits until each GPU is finished with its mini batch then takes the mean of all the gradients from the G GPUs and updates the model with this mean gradient.

Hybrid parallel optimization has several advantages and few disadvantages. As in asynchronous parallel optimization, hybrid parallel optimization allows for one to use multiple GPUs in parallel. Furthermore, unlike asynchronous parallel optimization, the incorrect gradient problem is not present here. In fact, hybrid parallel optimization performs as if one is working with a single mini-batch which is GG times the size of a mini-batch handled by a single GPU. However, hybrid parallel optimization is not perfect. If one GPU is slower than all the others in completing its mini-batch, all other GPUs will have to sit idle until this straggler finishes with its mini-batch. This hurts throughput. But, if all GPUs are of the same make and model, this problem should be minimized.

So, relatively speaking, hybrid parallel optimization seems to have more advantages and fewer disadvantages as compared to both asynchronous and synchronous optimization. For this report, we use the hybrid model.

2.2. Speech Recognition API

In recent years, chrome version 25 came up with the Web Speech API embedded in browser which support conversion of voice to text conversion in web applications. It is getting popular and going to be the future for voice recognition [10]. Browser exposes the speech recognition feature via the Speech Recognition interface. This interface has an ability to recognize voice context from an audio input (normally via the device's default speech recognition service) and respond appropriately. We have to create Speech Recognition object using JavaScript that has a number of event handlers available for detecting when speech is input through the device's microphone. We can also check the browser's compatibility using Webkit Speech Recognition present in browser window object. The Speech Grammar interface represents a container for a particular set of grammar that our app should recognize [11]. Grammar is defined using JSpeech Grammar Format.

2.3. Text to Voice Conversion

The output produced by the server will be in text format which need to be converted into voice. So, we need a solution to convert this text into voice in an efficient way. The voice response will provide an interactive feeling to the users. There are many options available in text to voice conversion, for example, many cloud hosted APIs, or standalone implementations using pre-trained models such as gTTS. Here we analysed two solutions such as gTTS and Web Speech API.

2.3.1. Text to Speech Engine Using Pre Trained Model

The text-to-speech (TTS) is the process of converting text data into a vocal audio form. The program takes an input text and using methods of natural language processing understands the linguistics of the language being used, and performs logical inference on the text. This processed text is passed into the next block where digital signal processing is performed on the processed text. Using many algorithms and transformations this processed text is finally converted into a speech format. This entire process involves the synthesizing of speech. We use Google's Text To Speech (gTTS) library for text to speech conversion. gTTS is a very easy to use python library

which convert text to audio file, which will be transferred to the client as blob data. This API support many languages include English, French, German, Hindi and many more.

2.3.2. Speech Synthesis API

Speech synthesis is an interface coming up with the browser Web Speech library API. Speech synthesis is accessed via the Speech Synthesis interface, a text-to-speech component that allows programs to read out their text content. It comes up various voice types, rate and pitch that we can configure in the synthesis voice [11].

2.4. CLOSED DOMAIN QUESTION ANSWERING (cdQA)

Closed Domain Question Answering (cdQA) is an NLP based Closed Domain Question Answering System. The mission of cdQA is to allow anyone to ask a question in natural language and get an answer without having to read the internal documents relevant to the question.

When we think about question answering systems, it appears as two different kinds of systems: open-domain QA (ODQA) systems and closed-domain QA (cdQA) systems. Open-domain systems deal with questions about nearly anything and can only rely on general ontologies and world knowledge. One example of such a system is DrQA, an ODQA developed by Facebook Research that uses a large base of articles from Wikipedia as its source of knowledge. As these documents are related to several different topics and subjects, we can understand why this system is considered an ODQA. Closed-domain systems deal with questions under a specific domain (for example, medicine or hospitality), and can exploit domain-specific knowledge by using a model that is fitted to a unique-domain database. The cdQA-suite was built to enable anyone who wants to build a closed-domain QA system easily.

The cdQA architecture is based on two main components: the Retriever and the Reader. When a question is sent to the system, the Retriever selects a list of documents in the database that are most likely to contain the answer. It is based on the same retriever of DrQA, which creates TF-IDF (term frequency-inverse document frequency) features based on uni-grams and bi-grams and compute the cosine similarity between the question sentence and each document of the database [11].

After selecting the most probable documents, the system divides each document into paragraphs and send them with the question to the Reader, which is basically a pre-trained Deep Learning model. The model used was the Pytorch version of the well-known NLP model BERT which was made available by HuggingFace. Then, the Reader outputs the most probable answer it can find in each paragraph. After the Reader, there is a final layer in the system that compares the answers by using an internal score function and outputs the most likely one according to the scores.

Here the pretrained model contains Bert Stanford Question Answering Dataset (SQuAD) which have 100,000+ question-answer pairs on 500+ articles. We can further train this model based on our domain and make it a closed cdQA system.

Among the above-mentioned different solutions, by considering various aspects such as accuracy, ease of use, maintainability and cost, the proposed system will be implemented with solution below. We use apache server with PHP as server side scripting language. Voice to Text using Speech recognition API. For text to voice conversion we use Speech Synthesis API. The question answering model use cdQA. The cdQA was hosted on Python Flask web server. The proposed system Architecture diagram is shown in Figure 1.

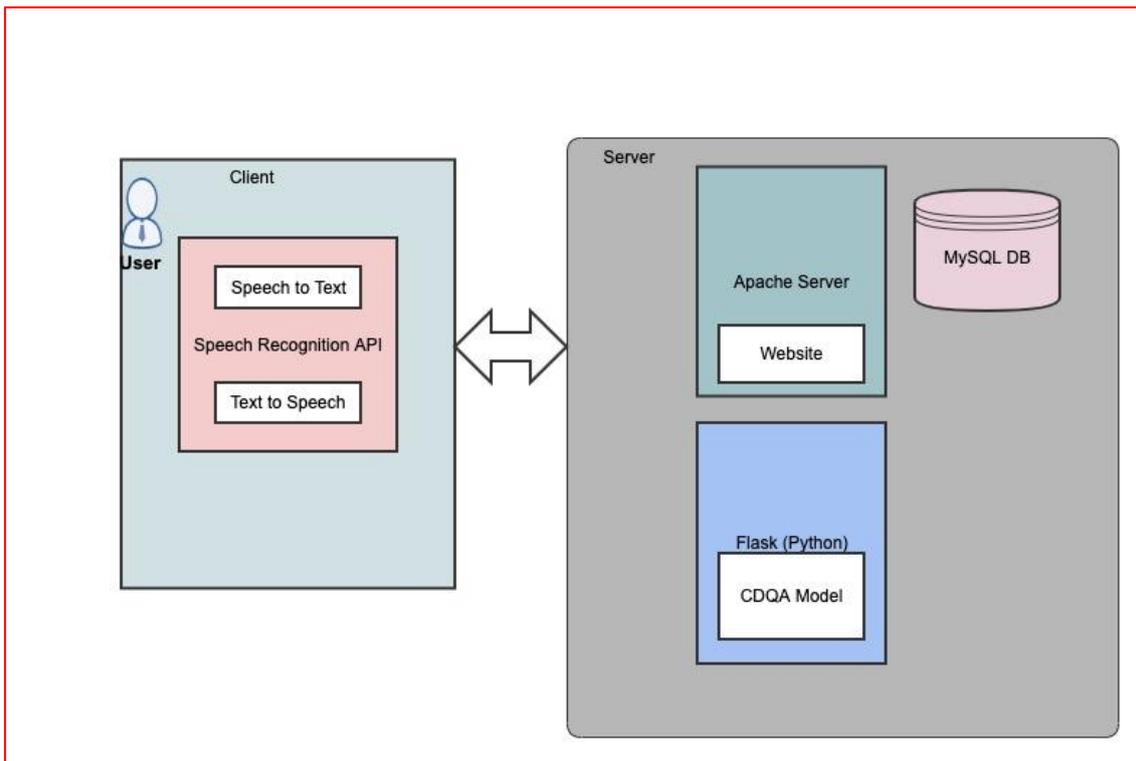


Figure 1. Proposed System Architecture

3. IMPLEMENTATION AND EXPERIMENTAL RESULTS

Based on the proposed system architecture the system has been implemented with the following. There are five views in this application. The UI shows main attractions nearby, major shopping areas, activities and details about the various kinds of rooms available. The website has been implemented using html, CSS, JavaScript, Bootstrap, FontAwesome for the client side, PHP and apache server for the server side and MySQL database as backend. Room availability is configured in database, it renders in UI when a user search for the rooms. At the top of the website, provided a button to initiate the voice chat. Home screen has been implemented as below, provided search feature at the left, where user can enter the date and guest number to search for the rooms. Click on search button will make rest API call to the server and load the available rooms from the database.

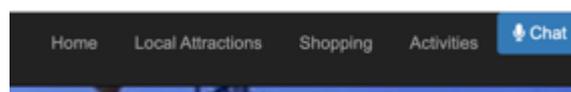


Figure 2. Home Page Menu

3.1. Closed Domain Question Answering (CDQA)

We retrained the model with hospitality pdf documents. Sample response for the question will show as below. The system not only shows outputs an answer, but also shows the paragraph where the answer was found and the title of the document / article.

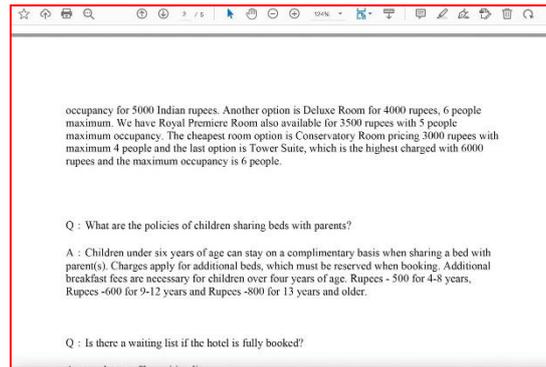


Figure 3. Sample cdQA model pdf

The model has been trained with 50-100 questions related to the hospitality. Each question-answer will be a paragraph. Once the model is trained, deployed, and exposed using Flask server. It exposes an endpoint which accept a string input and query the model and return JSON data which include answer to the query, paragraph where it found the answer and document name. Figure 3 shows a sample cdQA model pdf document.

3.2. Voice Chat Bot (EMMA)

The designed voice chat bot is named as “Emma”. Click on the “Chat (voice)” button at the top the navigation panel will show up a chat box as below with a welcome voice message as “My name is Emma, your voice assistance, how can I help you today”. This message is a configurable message.

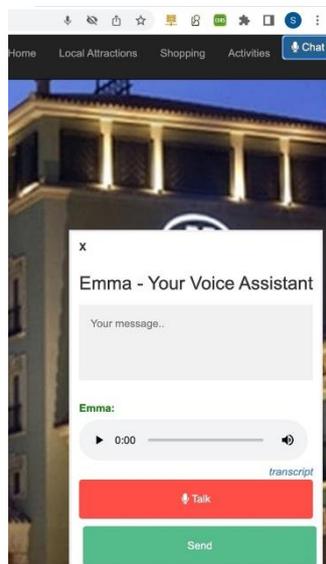


Figure 4. Voice chat bot screen

When user clicks on “Talk” button, it will start listening of voice input till the user stop talking. Using the Web Recognition API, it will convert the input voice into text and will show up in the text area which appears at the top of the voice chat popup as in Figure 5. Click on “send” will initiate a cdQA model hosted server call. The processing status will show up as the loading icon

on top left of the Talk button in Figure 5. Note the user's voice was transcribed into text and showed on the screen as well.

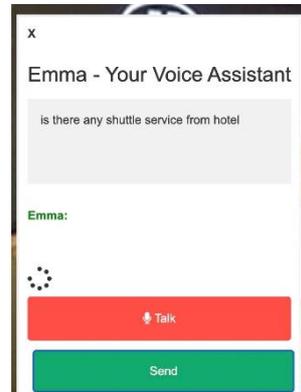


Figure 5. After user's voice input and the Send button clicked

Once the result is received, the audio response will render below the text area as shown in Figure 6. It also provides a transcript button, click on that will expand and show the text content.

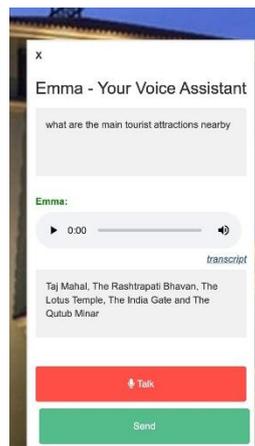


Figure 6: The voice and text response of the Chat bot

4. CONCLUSIONS AND FUTURE IMPROVEMENTS

Chatbots are software applications that use artificial intelligence and natural language processing to understand humans' need and guides them to their desired outcome. Here this project focused on implementing a voice based chatbot. The voice questions are parsed through an algorithm on a remote server that analyses the document for all possible relevant answers. The most relevant answer is sent back to the user, together with approximate confidence from the model. Reusable generic module which captures voice input and convert it into text and convert the text back to voice once the result received from the server. It is implemented using html and JavaScript library.

Based on the research on various solutions to implement voice to text and text conversion like Google Text-to-Speech (gTTS) engine and speech to text engine, the downside is both require separate infrastructure to maintain it. So decided to proceed with JavaScript utility wrap the implementation for speech recognition and speech synthesis API with event handling.

The question answering mechanism performed very well (based on the evaluation) and predictable for texts shorter than 3000 words. However, with longer texts it started to lose accuracy, losing track of details, and making significant mistakes. The summaries that were meant to help the user formulate questions also worked as intended, however, with the caveat that the summarization model exhibited unaccountable behaviour when supplied with longer texts. The actual usefulness provided are vary between documents. How to conquer this problem may require further research.

An incremental improvement on the system performance and accuracy can be achieved by retraining the model with more questionnaire and through a feedback loop. Training with the bigger dataset requires a higher end system with more GPUs.

REFERENCES

- [1] Brsuh, Kate and Scardina, Jesse. “Chatbot”, Techtargget. <https://searchcustomerexperience.techtargget.com/definition/chatbot>
- [2] Lishchynska, Daryna. “What Are Bots? How Do Chatbots Work?”, <https://botscrew.com/blog/what-are-bots/#:~:text=Chatbot%20or%20bot%20%E2%80%93%20is%20a,an%20instant%20pre%2Dset%20answer>
- [3] Soeyuenmez, Anil. “Customer Service Chatbots”, <https://www.messengerpeople.com/customer-service-chatbots-rule-based-vs-ai-what-you-need-to-know/>
- [4] Selig, Jay. “The Power Of Chatbots Explained”. <https://www.messengerpeople.com/customer-service-chatbots-rule-based-vs-ai-what-you-need-to-know/>
- [5] Dimitrios Buhalis and Iuliia Moldavska (2021). In-room Voice-Based AI Digital Assistants Transforming On-Site Hotel Services and Guests’ Experiences. In: Wörndl, W., Koo, C., Stienmetz, J.L. (eds) Information and Communication Technologies in Tourism 2021. Springer, Cham. https://doi.org/10.1007/978-3-030-65785-7_3
- [6] Bai Li, Nanyi Jiang, Joey Sham, Henry Shi, Hussein Fazal, “Real-World Conversational AI for the Hotel Bookings”. <https://arxiv.org/pdf/1908.10001.pdf>
- [7] Adam, M., Wessel, M. & Benlian, A. AI-based chatbots in customer service and their effects on user compliance. *Electron Markets* 31, 427–445 (2021). <https://doi.org/10.1007/s12525-020-00414-7>
- [8] Rajibul Hasan, Bernadett Koles, Mustafeed Zaman, Justin Paul, "The potential of chatbots in travel & tourism services in the context of social distancing", April 2021 *International Journal of Technology Intelligence and Planning* 13(1) DOI:10.1504/IJTIP.2021.10041470
- [9] “Welcome To Deep Speech Documentation”. <https://deepspeech.readthedocs.io/en/r0.9/>
- [10] Raun, Benson. “Voice to Text With Chrome Web Speech API”. <https://towardsdatascience.com/voice-to-text-with-chrome-web-speech-api-d98462cb0849>
- [11] Web Speech API <https://towardsdatascience.com/voice-to-text-with-chrome-web-speech-api-d98462cb0849>
- [12] “What Is A Voicebot?” Genesys. <https://www.genesys.com/en-sg/definitions/what-isa-voicebot>
- [13] Elupula, Vishnu. “How Do Chatbots Work? An Overview Of The Architecture Of Chatbots”, 15 May, 2019. <https://bigdata-madesimple.com/how-dochatbots-%20work-an-overview-of-the-architectureof-a-chatbot/>
- [14] The Ultimate guide to chatbot. <https://www.drift.com/learn/chatbot/>
- [15] Mazoor, Kashif. “What Are Chatbots? Beginners Guide To Chatbots”, July, 2017, <https://www.otechtalks.tv/wpcontent/uploads/2017/08/what-are-ChatBots.pdf>
- [16] Slesar, Mila. “Types Of Chatbots, An Overview For Business People”. <https://onixsystems.com/blog/types-of-chatbots-overview-forbusiness-people>
- [17] Wouters, Joren. “3 Chatbot Types, What Is Best For Your Business?” 20 December, 2020. <https://chatimize.com/chatbot-types/>
- [18] Lemaire, Adrein. “Hybrid Chatbot: How To Make Humans And Robots Work Together”. <https://www.dimelo.com/en/blog/integration-agentschatbots-botmind/>.
- [19] Gouba, Rubinder. “What Are Contextual Chatbots? How They Can Make A World Of Difference In User Experience?”, 6 October, 2018. <https://medium.com/makerobos/what-are-contextualchatbots-how-they-can-make-a-world-of-differencein-user-experience-e7446c96664e>

- [20] Engati, Team. Mapping The Growth Of Voice Enabled Chatbot!" <https://www.engati.com/blog/growth-of-voice-enabled-chatbots>
- [21] Expert.ai.Team. "What Is A Chatbot? Why Are Chatbots Important?", 17 March, 2020. <https://www.expert.ai/blog/chatbot/>
- [22] Farias, Andre. "How To Create Your Own Question Answering System Easily With Python", 7 July, 2019. <https://towardsdatascience.com/how-to-create-your-own-question-answering-system-easily-with-python-2ef8abc8eb5>
- [23] "Question Answering Using CDQA(BERT)". <https://www.kaggle.com/guizmo2000/question-answering-using-cdqa-bert-atos-big-data>
- [24] Bharath, K. "How To Get Started With Google Text-To-Speech Using Python", 26 August, 2020. <https://towardsdatascience.com/how-to-get-started-with-google-text-to-speech-using-python-485e43d1d544>

A COMPARISON BETWEEN VGG16 AND XCEPTION MODELS USED AS ENCODERS FOR IMAGE CAPTIONING

Asrar Almogbil, Amjad Alghamdi, Arwa Alsahli, Jawaher Alotaibi,
Razan Alajlan and Fadiyah Alghamdi

Department of Computer Science, college of Computer Science
and Information Technology, Imam Abdulrahman Bin Faisal University,
Dammam, Saudi Arabia

ABSTRACT

Image captioning is an intriguing topic in Natural Language Processing (NLP) and Computer Vision (CV). The present state of image captioning models allows it to be utilized for valuable tasks, but it demands a lot of computational power and storage memory space. Despite this problem's importance, only a few studies have looked into models' comparison in order to prepare them for use on mobile devices. Furthermore, most of these studies focus on the decoder part in an encoder-decoder architecture, usually the encoder takes up the majority of the space. This study provides a brief overview of image captioning advancements over the last five years and illustrate the prevalent techniques in image captioning and summarize the results. This research study also discussed the commonly used models, the VGG16 and Xception, while using the Long short-term memory (LSTM) for the text generation. Further, the study was conducted on the Flickr8k dataset.

KEYWORDS

Image Captioning, Encoder-Decoder Framework, VGG16, Xception, LSTM.

1. INTRODUCTION

One of the most challenging and important topics in computer vision and natural language processing is image captioning [1], [2]. Image captioning aims to generate a natural language description based on the association between the objects in the given image. Image captioning can be helpful in different applications such as human-computer interaction and providing help for visually impaired persons [3]. Therefore, several studies have developed an image captioning model [4,5]. Initially, the studies related to image captioning were focused mainly on generating natural language descriptions for video [6], following the studies describing neural caption generation architectures [7, 8], such as the encoder-decoder architectures proposed in [9]. Recently, the encoder-decode architecture has shown much improved outcomes in efficiently generating natural language descriptions of an image [10]. At first, the CNN layers are used to extract the features of the image. Then the collected features are used by the Recurrent neural network (RNN) model to attain the information from the image [11].

This study reviews the current advancement of image captioning models and summarizes the underlying framework. Although much attention has been paid to the decoder, there has not been enough focus on the encoder. To fill this gap, this study will compare the performance of two

different encoder models, namely: VGG16 and Xception. Moreover, a comparison that focuses mainly on the performance of two widely used encoders - VGG16 and Xception is poorly investigated, which will help further researchers to decide on the encoder model.

The rest of the paper is organized as follows. Section 2 presents related work. Section 3 discusses the materials and methods used in this work. Experiments done are described in Section 4. The result obtained is illustrated in section 5. Conclusions and future work in Section 6.

2. RELATED WORK

In this section, we will summarize multiple related studies from different sources. The studies will be organized in chronological order ascendingly. The purpose of the related work is to gain an understanding of the published studies relevant to the image captioning field.

In [12], they used the MSCOCO dataset and LSTM to encode the text and used CNN as an image encoder to extract features, and they obtained the best result compared with their benchmark. Another study [13] used VGG16 as an encoder, which aids in creating image encodings. Then, the encoded images are fed into an LSTM. The proposed model was enhanced with hyper-modifying parameters. As a result, the model's accuracy increased to attain state-of-the-art results. In [14], different models of image captioning were used. A merge architecture was applied in this study. CNN-5, vgg16, and vgg19 are the different CNN that are used along with the LSTM. The experiment is done on Flickr8K dataset. A Bilingual Evaluation Understudy (BLEU) evaluation metric is used to evaluate the models. The result showed that VGG16 performs better than other models. The authors in [15] compared different models of image captioning. All models were conducted on the Flickr8K dataset. The architecture used in this study is encoder-decoder architecture. For the encoder, two different CNN models are used, which are VGG16 and InceptionV3. For the decoder part, two types of LSTM were used. The first type is a unidirectional LSTM that works in one direction. The second type is bidirectional LSTM which works in two directions. The proposed models used greedy search and beam search algorithms to generate the captions. The results show that the InceptionV3 with bidirectional LSTM with beam search gave the best result. The evaluation metric used is BLEU. In [16], the study proposed an image caption generator in the Bengali language using a merged dataset of two languages by combining flickr8k, BanglaLkey, and Bornon datasets. The transform-based and visual attention approaches were used to implement the proposed model. The Transform-based approach used an inceptionV3 encoder and fed to a dense layer that contains an activation function. The visual attention approach implements an Encoder-decoder framework as well. In the encoder part, the InceptionV3 and Xception models were used. For the evaluation of the proposed model, the BLEU and Metor were used.

In [17], the study proposed an image captioning model to use the model on any website to generate the description of the inputted image. The proposed model followed the CNN-LSTM concepts and was conducted on the flicker8k dataset. In [18], the study used CNN and RNN models, and the Xception was trained using the flicker8k dataset. Another study used the xception model coupled with LSTM in [19] to discover the object found in the image, detect the relationship among the objects, and generate the proper captions. This study was trained using the flicker8k dataset. The criteria to evaluate the model was the loss value. In [20], the authors compared the most popular CNN architecture: Xception, Resnet50, InceptionV3, Vgg16, and Densent201. Along with the LSTM decoder. The comparison was done to see the effect of the performance by implementing different encoder models. The study used flicker8K dataset. The evaluation of the comparison was the loss value and the accuracy to compare the model's performance. The study [21] proposed different CNN models VGG16, Xception, and inception coupled with bi-directional layer RNN models for an enhanced image captioning model. The

models were trained using flicker30K and coco datasets. The BLUE score and training and loss are used to evaluate each model.

3. MATERIALS AND METHODS

This section includes the description of the dataset used in the study and the different encoders: VGG16 and Xception. Finally, the decoder model.

3.1. Dataset pre-processing

The dataset used in this work is Flickr8k, and it is available on GitHub [22]. Flickr8k consists of two folders, the first folder contains only images, and the second folder contains a text file with the image descriptions. For the data pre-processing phase, we start working on the text file and organize it by mapping the image ID to a list of five corresponding descriptions. After that, we worked on data cleaning by making all letters in lower case, removing all the punctuations, and removing words with one character (e.g. 'A'). Lastly, we saved all changes made in a new text file.

3.2. The Encoder models

3.2.1. VGG16 model

VGG16 is one of the most preferred CNN models as it has a very uniform architecture. Simonyan and Zisserman developed this model in 2014 [23]. It contains 16 convolutional layers. By having this amount of layers, the complexity would increase compared to the initial versions of the CNN architecture. In the below Figures, the size is proportionally getting reduced. The two layers are convolutional, and the output of these two layers is 224x224, followed by the max-pooling layer, and the final output after the max-pooling layer of size 2x2 and stride of 2 will be reduced to 112x112. Finally, we have three fully connected layers called dense. Figure 1 shows the architecture of the VGG16 model.

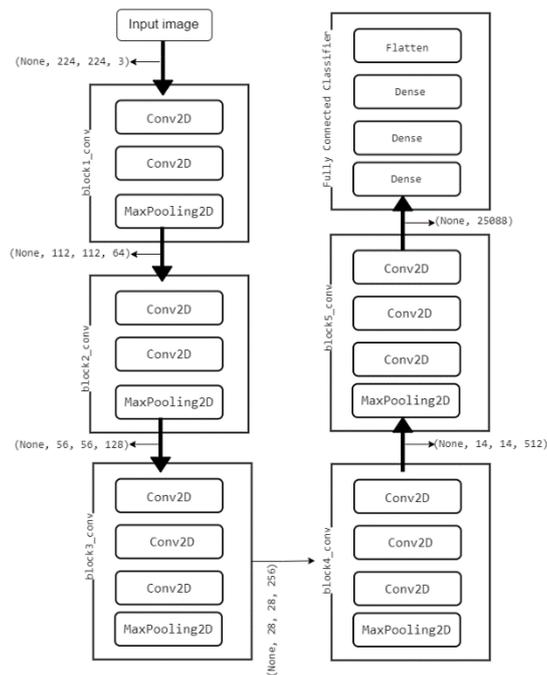


Figure 1. VGG16 Architecture

3.2.2. Xception model

The Xception model, also called “Extreme Inception” was proposed by Francois Chollet. It is a kind of CNN model used to extract the features from the image. Also, it is an extension of the inception model that is also considered a type of CNN model [24], but a better and enhanced version by reversing some steps to be more efficient and easier to modify [25]. The Xception model contains 37 layers [20]. The model uses the depthwise separable convolutional layers approach, which divides the image into K input channel with depth equal to 1, then applies the filter into each part with depth equal to 1, after that compressed all input channels space then applying 1*1 convolutional. The accuracy of the Xception model considers the highest among the CNN model in agreement with the LR in [15]. Therefore, it gives the best result compared to the other CNN models. Figure 2 illustrate the layers of the Xception model.

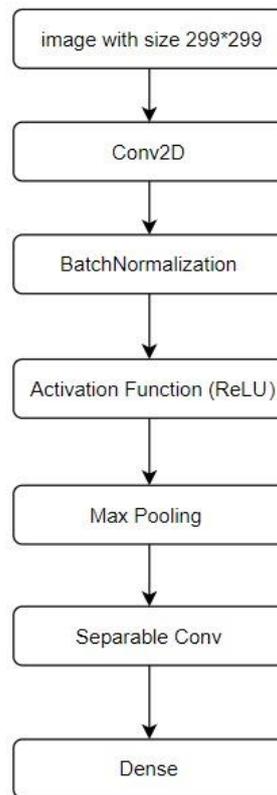


Figure 2. Layers of Xception Model

3.3. The Decoder model

For the decoder model, LSTM based model was used, which takes input from the feature extraction model to predict a sequence of words, called the caption.

Because LSTM overcomes the RNN's constraints, LSTM is more effective and superior to the regular RNN. With a forget gate, LSTM can keep relevant information throughout the processing of inputs while discarding non-relevant information. It can process not only single data points but also complete data sequences [26].

4. EXPERIMENTAL

For the experiments, our model follows the encoder-decoder framework. Therefore, we tested and evaluated two different encoder models. Furthermore, we illustrated the conducted processes for developing the models for each model and how we trained the models. Whereas the decoder remains fixed during the experiment, as mentioned before, in order to focus on comparing the performance of the encoder model.

4.1. The encoder

In the feature extraction step, the size of the image features is 224x224. Extracting the features of the image is done before the last layer. The goal of the last layer is to predict the classification of an image. For this reason, the last layer is dropped. The models were trained on Flickr8k dataset as was described in Section 3.

4.1.1. VGG16

• Before optimization

When we started the model's training, we split the dataset into two parts. The first part is for training, and the second part is for testing. Flickr8k dataset contains a file named "Flickr_8k.trainImages.txt" that includes 6000 image ID; this file is used for the training part. The training phase will be done in three steps. The first step, load the features extracted from the VGG16 model. In the second step, we will initiate a dictionary that contains descriptions for each image. The third step, create tokenizing vocabulary by using Keras, which provides the tokenizer class, and it can do the mapping from the loaded description data. In this step, we need to fit the tokenizer given the loaded photo description text. The `create_tokenizer()` function is responsible for fitting the created tokenizer given the loaded photo description text. In addition, it's for mapping each word of vocabulary with a unique index value.

• After optimization

To optimize the result and reduce the loss obtained, we implement Adam algorithm, which is an optimizer that increase efficiency of neural network weights.

4.1.2. Xception

• Before optimization

Our CNN-RNN model consists of three main parts: feature extraction (encoder), sequence processor, and decoder. In the experiment, we used images with a size equal to 299x299. In the features extraction step, which is done before the last layer of the model, we got an 8091 feature vector. In training, feature extraction is loaded to the model, and the dataset is divided into two parts: training with 7091 images and testing with 1000 images. Then, we tokenized the vocabulary by mapping each word with a unique index value, and each image will have a maximum length of sentence equal to 31. After that, we created a data generator to train the model to yield the image in batches.

• After optimization

The Adam algorithm was implemented to optimize the model to improve its performance.

5. RESULT AND DISCUSSION

In this study, a total of four models were tested and evaluated —VGG16, VGG16 with optimization, Xception, and Xception with optimization. The criteria for the comparison are taken to be the loss instead of the accuracy value, and the standard metric for comparison used here is the BLEU score.

Table 1. Evaluation Table

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
VGG16 Epoch= 100 Loss=3.0345	0.522997	0.279958	0.186401	0.079141
VGG16 with optimization Epoch = 100 Loss= 3.3746 Optimizer= Adam	0.498937	0.251331	0.168155	0.068864
Xception Epoch= 50 Loss= 4.3955	0.096406	0.031889	0.020180	0.004638
Xception with optimization Epoch= 50 Loss=3.3618 Optimizer= Adam	0.550791	0.309441	0.216791	0.105341

The above table shows each model's performance in terms of the BLEU score, testing loss of the implemented models, and the number of epoch with the optimizer if used.

Our results demonstrate that Xception with optimization BLEU scores outperformed the other three models. The highest BLEU score achieved in the study was 0.550791. Both Xception with optimization and VGG16 before optimization have similar scores. However, the loss of VGG16 was less than Xception with optimization. The main motivation for using the adam algorithm was to show a significant improvement in the runtime and memory consumption and increase the efficiency of neural network weights, as mentioned in the previous section. The caption generated from the Xception with optimization model gives the best probability and more accurate captions (see Figure 6). In contrast, the captions generated by the other three models (Figure 3-5) were long sentences compared to Xception with optimization. We can infer from the experiment that when the sentences are long, the more probable to make mistakes. In most situations, we found that the short sentences are sufficient to explain an image, whereas lengthier sentences frequently contain duplicate information and grammatical errors. The main challenge was to reduce the loss in Xception models, and after using the optimizer, the loss decreased. Yet, it remained higher than the loss obtained in VGG16 before optimization (see figure 7). Hence, we observed that when the number of the Epoch is increased, the number of loss models will increase in the Xception models due to the small size of the dataset.



startseq two men are playing soccer on field endseq

Figure 3. VGG16 Before Optimization



startseq man in black shirt and black pants is sitting on the street endseq

Figure 4. VGG16 After Optimization



the man is sitting on the top of the rock

Figure 5. Xception Before Optimization



startseq dog is running through the grass endseq

Figure 6. Xception After Optimization

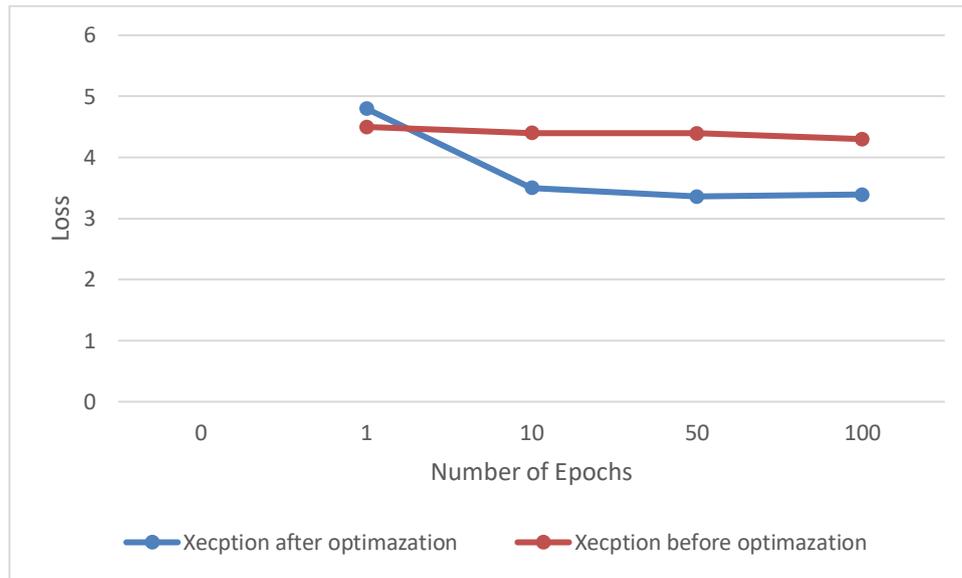


Figure 7. Testing Loss Curve for Xception Before and After Optimization.

6. CONCLUSION

In this study, we used an encoder-decoder framework that been used in the previous studies. We evaluated two different encoder models for the purpose of comparing the VGG16 and Xception encoder models. So far, no study has been published comparing these two models which will help researchers figure out which model is outperforming the other. The outcome of the comparison shows that the Xception model, when implemented adam algorithm, will generate the most accurate caption compared to the other three models. Moreover, the study attempted to use Flickr8k open-source datasets. Despite the precise caption achieved, there is still a need for a larger dataset. A large dataset will enhance the model's performance.

ACKNOWLEDGEMENTS

We would like to thank Ms. Asrar Almogbil for her cooperation on providing the instructions. We also extend our appreciation to Dr. Nida Aslam and Ms. Abrar Alotaibi for their continuous efforts in helping and answering our questions during the experiment.

REFERENCES

- [1] Raimonda Staniūtė and Dmitrij Šešok. A systematic literature review on image captioning. *Applied Sciences*, 9(10):2024, 2019.
- [2] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*, pages 2425–2433, 2015.
- [3] Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, Jose MF Moura, Devi Parikh, and Dhruv Batra. Visual dialog. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 326–335, 2017.
- [4] A. Ramisa, F. Yan, F. Moreno-Noguer, and K. Mikolajczyk, “Breakingnews: Article annotation by image and text processing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1072–1085, 2018.

- [5] H. Ben, Y. Pan, Y. Li et al., "Unpaired image captioning with semantic-constrained self-learning," *IEEE Transactions on Multimedia*, vol. 24, pp. 904–916, 2021.
- [6] Vinyals, Oriol, et al. "Show and tell: A neural image caption generator." *Computer Vision and Pattern Recognition (CVPR)*, 2015 IEEE Conference on. IEEE, 2015.
- [7] M. Tanti, A. Gatt, and K. P. Camilleri, "What is the Role of Recurrent Neural Networks (RNNs) in an Image Caption Generator?," Aug. 2017.
- [8] Sulabh Katiyar and Samir Kumar Borgohain. Comparative evaluation of cnn architectures for image caption generation. *arXiv preprint arXiv:2102.11506*, 2021.
- [9] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." In *Advances in neural information processing systems*, pp. 3104-3112. 2014.
- [10] F. Huang, X. Zhang, Z. Zhao, and Z. Li, "Bi-directional spatial-semantic attention networks for image-text matching," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2008–2020, 2019.
- [11] S. Li, Z. Tao, K. Li, and Y. Fu, "Visual to text: Survey of image and video captioning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 4, pp. 297–312, 2019.
- [12] A. T. S. B. a. D. E. Oriol Vinyals, "Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge," *IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 39, p. 12, 2017
- [13] V. V. P. M. M. Ashish Pateria, "Enhanced Image Capturing using CNN," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 4, p. 6, 2019.
- [14] A. a. D. S. a. Y. M. V. Jmail, "IMAGE CAPTIONING: TRANSFORMING SIGHT INTO SCENE," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 02, no. 06, pp. 54-66, 2020.
- [15] S. Takkar, A. Jain, and P. Adlakha, "Comparative Study of Different Image Captioning Models." *Fifth International Conference on Computing Methodologies and Communication, India*, 2021.
- [16] F. M. Shah, M. Humaira, M. A. R. K. Jim, A. S. Ami and S. Paul, "BORNON: BENGALI IMAGE CAPTIONING WITH TRANSFORMER-BASED DEEP LEARNING APPROACH," *arXiv*, p. 20, 2021.
- [17] M. M. Patil, "Experiment based on Deep Learning: Image," *INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS - IJCRT* , vol. 9, no. 12, p. 6, 2021.
- [18] N. L. C. K. Satyabrat Mandal, "Automatic Image Caption Generation System," *International Journal of Innovative Science and Research Technology*, vol. 6, no. 6, p. 4, 2021.
- [19] V. U. G. S. V. M. Megha J Panicker, "Image Caption Generator," 2021.
- [20] S. R. Sahrial Alam, "Comparison of Different CNN Model used as Encoders for Image Captioning," 2021.
- [21] A. P. Yash Indulkar, "Comparative Study for Neural Image Caption Generation Using Different Transfer Learning Along with Diverse Beam Search & Bi-Directional RNN," 2021.
- [22] The dataset available in: https://github.com/goodwillyoga/Flickr8k_dataset
- [23] A. D. Hussam, "Compressed residual-VGG16 CNN model for big data places image recognition," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, 2018.
- [24] M. j. Panicke, V. Upadhayay, G. Sethi and . V. Mathur, "Image Caption Generator," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 10, no. 3, p. 6, 2021.
- [25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [26] P. G. Shambharkar, P. Kumari, P. Yadav, and R. Kumar, "Generating Caption for Image using Beam Search and Analyzation with Unsupervised Image Captioning Algorithm." *Fifth International Conference on Intelligent Computing and Control Systems, India*, 2021.

OUTLIER DETECTION AND RECONSTRUCTION OF LOST LAND SURFACE TEMPERATURE DATA IN REMOTE SENSING

Muhammad Yasir Adnan, Prof Yong Xue and Richard Self

School of Computing & Engineering, University of Derby, United Kingdom

ABSTRACT

In quantitative remote sensing, missing values classified as outliers occur frequently. This is due to technical constraints and the impact of weather on the efficiency of instruments to collect data. In order to deal with these missing values, we offer an Outlier-Search-and-Replace (OSR) algorithm that uses spatial and temporal information for the detection and reconstruction of missing data. The algorithm searches for outlier in the data and reconstruct by finding the best possible match in spatial locations.

KEYWORDS

Remote Sensing, Missing Data Reconstruction, Outlier, MODIS, Land Surface Temperature.

1. INTRODUCTION

The temperature of the terrestrial surface is an essential indicator of the state of the atmosphere. This is extensively used in a wide range of environmental applications, including agriculture. Conditions in the atmosphere have a significant impact on the ability of remote sensing sensors to gather information. For gathering information on the atmosphere, ocean, and land surface, these instruments are the most often used way of data collection. Outlier Search and Replace (OSR) is a technique for detecting and reconstructing outliers in land surface temperature data that is presented in this article. Land surface temperature data from the Moderate Resolution Imaging Spectroradiometer (MODIS) collected in January 2018 is being used for the experiments. The results show that the suggested approach, which takes advantage of both spatial and temporal information, works well when it comes to detecting and reconstructing missing land surface temperature information.

The remainder of the paper is arranged in the following manner. Section 2 discusses the work that is related to it. Section 3 of this work provides a detailed discussion of the outlier identification and reconstruction method that has been proposed in this study. Section 4 summarises the results of experiments conducted using the OSR model. In the end, Section 5 presents the conclusion.

2. RELATED WORKS

The remote sensing platforms are comprised of the equipment or vehicles that are used to collect data from the field. The earth observation data is highly complicated and susceptible to inaccuracies due to the way it is collected [1]. One of the characteristics of the sensors installed is the time of image accusation, the distance between the object and the sensor, the interval between image accusation and image location, and the range of coverage. The missing values are classified as

outliers in this study. There are many different types of missing information, which can be broadly classified into the following categories:

- Sensor Malfunction
- Cloud Obscuration

Anomalies introduce outlier in the form of malicious data which is inconsistent with respect to rest of the data. The Figure 1 demonstrates the presence of outliers.

Sensors play a vital role in remote sensing information gathering. The failure of which leads to missing information. For example 15 of 20 detectors in MODIS Aqua band 6 give malicious readings [2]. Typical example of missing data recovery caused by sensor failure is presented by [3]. Sensor malfunction leads to the phenomenon of striping in remote sensing images. An example of striping in a remote sensing images is shown in Figure 2 [4].

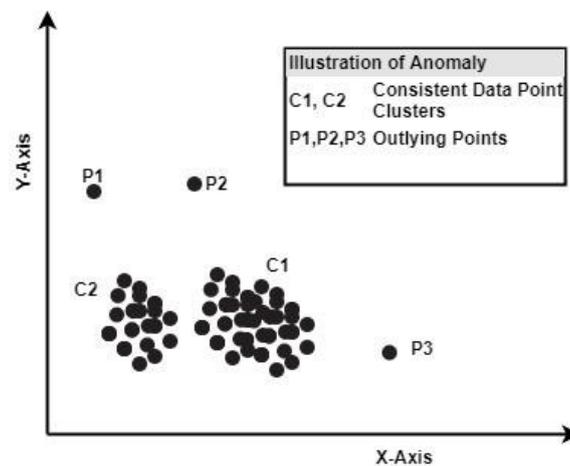


Figure 1. Illustration of Anomaly



Figure 2. Striping in Remote sensing imagery

Cloud covers hinders the information captured in remote sensing images which greatly reduces the data usability in subsequent application. At one time, 35% of earth surface is covered by clouds [5] and if an individual country is considered, e.g Canada has 50% to 80% of its land is covered with cloud in the morning [3]. Landsat ETM+ scenes are 35% contaminated by clouds which amounts to significant loss of data [6]

Detection of outlier and reconstruction is mainly classified into four categories [3, 5, 7]

- Spatial Based Methods
- Spectral Based Methods

- Temporal Based Methods
- Hybrid Methods

Image inpainting is the essence of spatial based methods. These are many traditional methods of image reconstruction used in the area of remote sensing and computer vision. In image inpainting, it usually assumes the fact that the missing information shares similar geographical features and fills the information gaps with this idea [8]. No Auxiliary images are required for spatial methods [9]. Spatial methods follows the correlation between local and global information of the image to fill the information gap but as the reference data is not large, spatial methods are mainly suitable for reconstruction of small missing areas and results are not guaranteed for complex terrain.

Spectral bands in remote sensing imagery are correlated and the information from another band can be utilized to reconstruct the missing information which overcomes the lack of prior information problem in spatial based methods. When hyper spectral or multi-spectral images have missing information, they both have bands with complete and missing information so the idea is to utilize the bands with complete information to reconstruct missing information by establishing a correlation between the bands. Spectral methods also known as multi-spectral-complementation methods [10]. For example Aqua MODIS has repeated patterns of black strips in its imagery due to sensor malfunction in band 6. The solution to this problem was first proposed by [2] in which author states that Aqua MODIS band 6 and 7 are correlated with coefficient of 0.9821 and 0.9777. Therefore missing information in band 6 can be recovered from highly correlated band 7.

Thick cloud cover causes all the spectral bands to be contaminated and have missing information in them and also the defective sensors may cause missing information in all the bands for a particular spot. Therefore spectral methods become useless as they are based on spectral correlation which is destroyed after all the bands have missing information. At this point, temporal based methods come into play. As the clouds are continuously moving and data can be acquired for same region at some other time interval. Determination of time interval is a tricky part in this method as if time interval is large, it will be affected by land cover change, but if land cover is small it will have overlapping clouds in two time slots. Lots of work has been done by researchers in temporal based methods, for example [11] presented a method using local linear histogram matching (LLHM) which required high quality data to function but it ends up giving poor results for heterogeneous landscape [12–14] proposed algorithms to improve radiometric consistency of multi-temporal images for heterogeneous land.

Temporal methods outperform all other methods but due to the limitation of amount of land cover change, temporal methods under-perform. Spatio-temporal fusion based method use data fusion from multiple sources to overcome this limitation. Instead of using one reference image [15] used two reference image in close dates to the target cloudy image. The errors produced by temporal methods due to significant land cover changes are avoided by this idea. This method further uses a residual correction strategy to improve spectral similarity between recovered area and remaining cloud free region. Spatial, temporal and spectral method discussed earlier have

their own strength and weaknesses, STS methods are developed by considering the strengths of all these methods to reconstruct missing information much more efficiently and accurately. [4] came up with a unified model called as spatial-temporal-spectral (STS) model based on deep convolutional neural network. The model not only solve the problem of dead lines in MODIS band 6 but also solves corrector-off problem in Landsat Enhanced Thematic Mapper imaging. It is also able to remove thick clouds and shadows using multi-source data. The method establishes a mapping between missing data and complete data with auxiliary data using a deep CNN. The model uses a residual output to learn the relation between different auxiliary data. These methods are also known as hybrid methods [3]. Missing LST data reconstruction for clear sky conditions can overestimate as compare to the reconstructed data under cloudy conditions [16]. While there is a limitation, there is till enough research gap to produce high quality reconstructed land surface temperature data. The author in [17] proposed a robust gap filling method by fusing MODIS and VIIRS LST data. Most algorithms presented in the literature uses one auxiliary image for reconstruction of data. Our algorithm uses auxiliary information from multiple sources as well as multiple days to detect and reconstruct missing values.

3. MODEL

3.1. Dataset

MODIS LST provides observations for daytime and nighttime. These images are taken from MYD11A1 and MOD11A1 which are MODIS LST products. MOD11A1 captures data at 10:30AM and PM while MYD11A1 captures data at 01:30AM and PM. The image shown in Figure 3 shows data from January 1st 2018 to January 5th 2018 from both the MODIS products. There is high correlation between the data from two MODIS products [18] which indicates that temporal information from similar or multiple satellite product can be used effectively for detection and reconstruction of missing data. The raw data from MYD11 and MOD11A1 is downloaded (<https://modis.gsfc.nasa.gov/data/dataproduct/>) and true color images are obtained by processing in ArcGIS with yellow color depicting the missing values.

3.2. Study Area

The dataset is collected from Moderate Resolution Imaging Spectroradiometer (MODIS) and Beijing–Tianjin–Hebei Figure 4 region is selected for initial experimentation. The target area has been used on numerous occasion for various remote sensing applications. The area is approximately 218000km² and is located close to the Northwest Pacific Ocean.

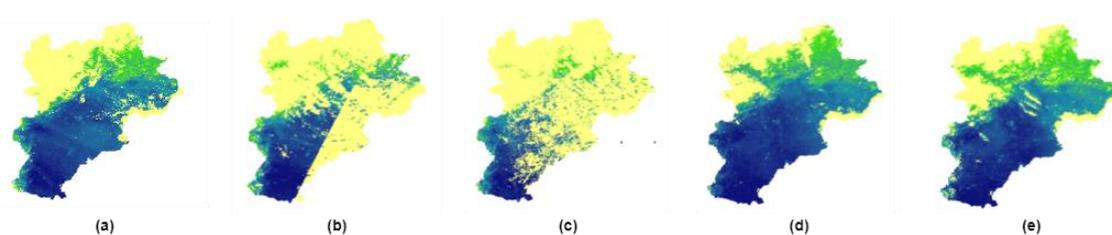


Figure 3. MODIS Input Image Series of 4 Days

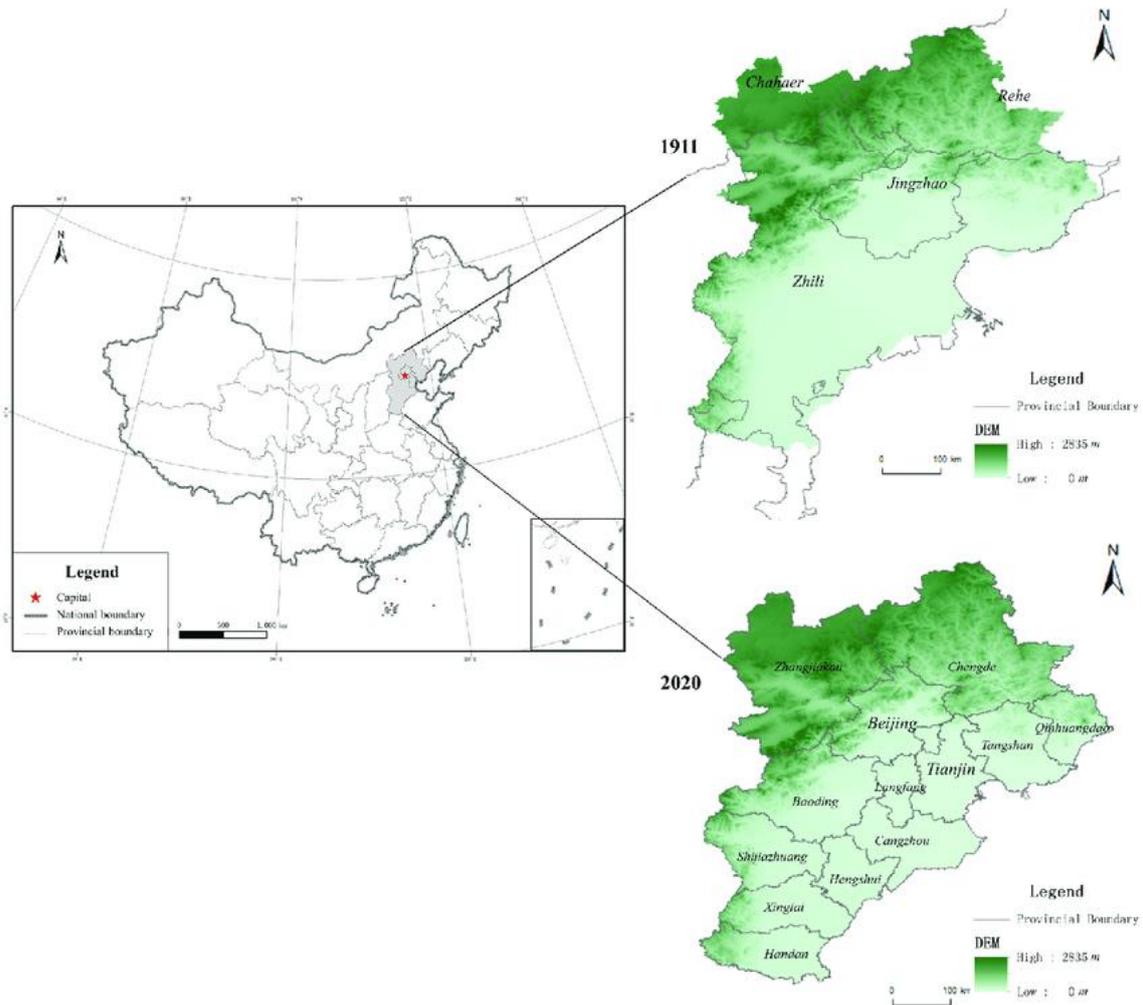


Figure 4. Beijing-Tianjin-Hebei Spatial Location [19]

3.3. Outlier Detection

Region of interest (ROI) of size MXN is selected which contains missing values to be detected and reconstructed. Once ROI_w is selected, the same region of interest is selected on all the input images to form a mosaic of temporal information as shown in



Figure 5. Image MOSAIC of $M \times N$ windows

Each image slice represent data from one day. Data for each day could be from a different satellite as well with similar temporal and spatial information which makes this algorithm efficient in order to detect and reconstruct missing values.

Dynamic time warping compares each pixel of image mosaic from ROI_w using Equation (1) to form a series of distance values.

$$C_d, D_d = DTW (I_{iP(j,k)}, I_{i+1P(j,k)},) \quad (1)$$

$P(j, k)$ is the pixel value at location j, k in each successive day input image region of interest ROI_w . D_d holds distance values between pixels of each successive image slice and C_d holds the coordinate of each pixel being compared. As there is high correlation between the temporal LST values of same region in successive days as well as spatial information of data from same satellite as well as multiple satellite [17]. A linear distance curve is formed when pixel values are correct but whenever there is a missing value in the image, the distance value is very high indicating anomaly. These values are identified and located as outliers based on a threshold value T_h and reconstructed in the next section. The threshold value T_h is currently being obtained by experimentation by comparing true pixels and known outliers.

3.4. Missing Data Reconstruction

Once the location of the outlying values C_{d_o} are identified. The reconstruction process begins. The algorithms traverses through the list of distance values D_{d_o} at outlier location C_{d_o} in all temporal image slices in mosaic and finds a pixel value with least distance value using Equation (2).

$$D_d = Min(C_D, D_D) \text{ for } 1 \leq R \leq N \quad (2)$$

C_D in Equation (2) gives the location of the pixel P_{C_D} with lowest distance D_d between pixel of image slices in mosaic at location similar to outlier C_{d_o} . The pixel P_{C_D} is taken as reference pixel to reconstruct the outlier.

Now a 3×3 window W_o is taken around the location of outlier and the reference pixel P_{C_D} is compared using Equation (2) with the neighboring pixels of the outlying value in W_o . The pixel which is at the lowest distance from the reference pixel is copied at the outlier location. This algorithm is similar to spatial reconstruction of missing values but in this case temporal information is also being utilized to identify the outliers.

4. RESULTS

The images in Figure 3 are used as input to test the accuracy of the algorithm. Outliers were introduced randomly in the input window of $M \times N$ Once the reconstruction process is complete, the reconstructed values are compared with original values before the outliers were introduced. The results are shown in Table 1. The matches show the number of pixels whose values were remained same after the reconstruction and non matches shows number of pixels whose values were different after the reconstruction process which algorithm is finding the best possible match by looking at spatial information in the image.

Table 1. Window Size and number of Random outlying Pixels

Window Size	Number of Random Outliers	Matches	Non Matches
4 X 4	5	1	5
6 X 6	10	0	10
8 X 8	15	7	8
10 X 10	32	24	9
15 X 15	62	21	41

The actual image and reconstructed image were compared using Equation (1) and based on matches and non matches and average distance between each of the reconstructed and original pixel value is shown in Figure 6. The average distance was calculated by adding the distance values between actual pixels and reconstructed pixels and dividing by number of respective outliers for window sizes.

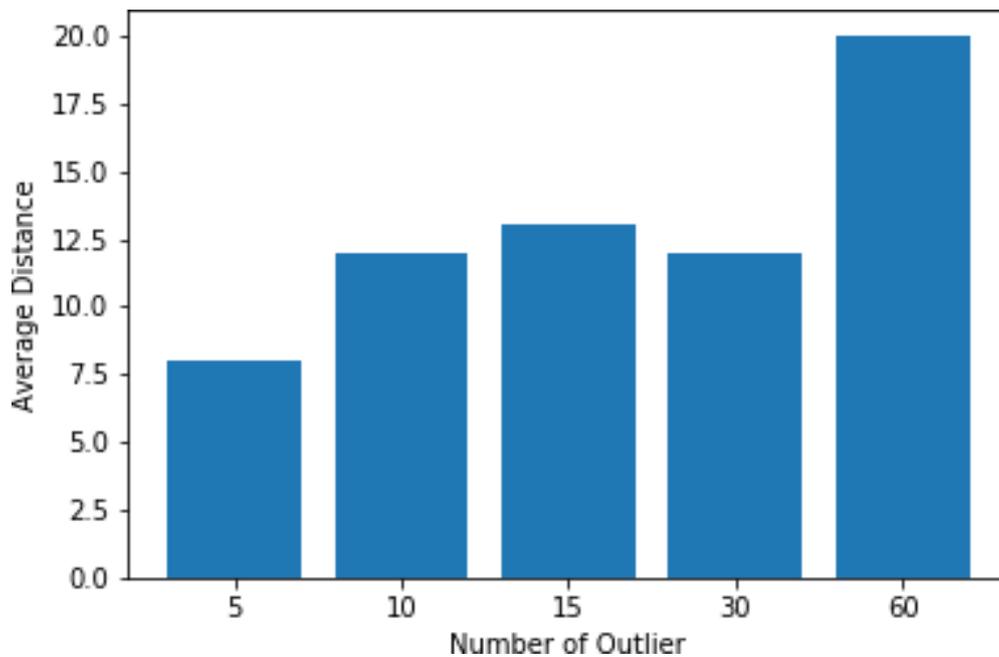


Figure 6. Average Distance between Original and Reconstructed value

The Figure 7 shows pair of image from a-b with input image with outliers and reconstructed image.

5. CONCLUSION

The proposed algorithm for outlier detection and reconstruction in remote sensing makes use of both spatial and temporal information, and as a result, it is referred to as a hybrid algorithm of outlier detection and reconstruction in remote sensing. In order to maximise efficiency in terms of exploiting auxiliary information, the algorithm can make use of temporal information from similar resources as well as from numerous resources at the same time which makes it very flexible.

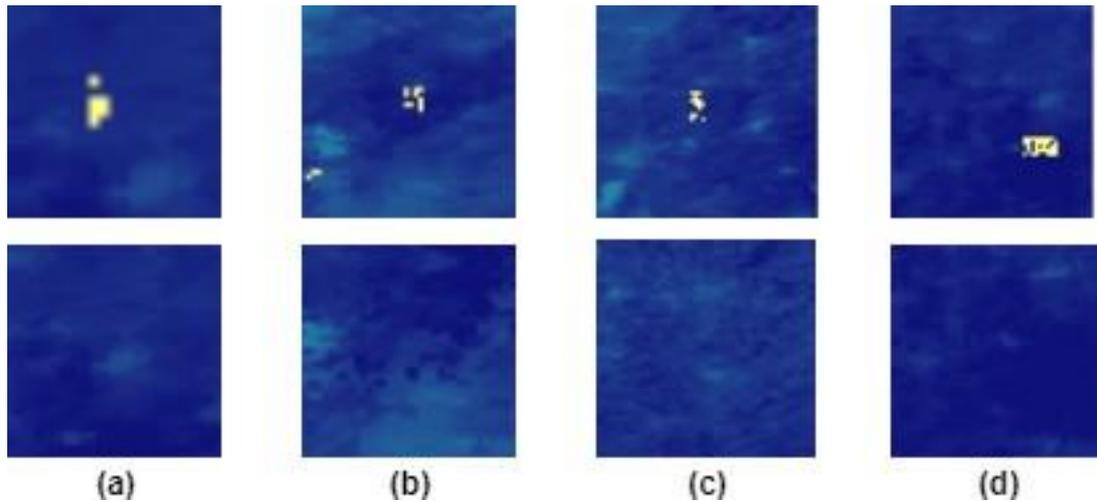


Figure 7. Outlier and Reconstructed Image Pair

REFERENCES

- [1] Alexandria Dominique Farias and Gongling Sun, (2020) “Data Mining and Machine Learning in Earth Observation-An Application for Tracking Historical Algal Blooms”, in *CS & IT Conference Proceedings*, Vol. 10.
- [2] Lingli Wang, John Qu, Xiaoxiong Xiong, Xianjun Hao, Yong Xie, and Nianzeng Che, (05 2006) “A new method for retrieving band 6 of Aqua MODIS”, *Geoscience and Remote Sensing Letters, IEEE*, Vol. 3, pp. 267–270.
- [3] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, (9 2015) “Missing Information Reconstruction of Remote Sensing Data: A Technical Review”, *IEEE Geoscience and Remote Sensing Magazine*, Vol. 3, No. 3, pp. 61–85.
- [4] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, (8 2018) “Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 8, pp. 4274–4288.
- [5] C. Lin, K. Lai, Z. Chen, and J. Chen, (1 2014) “Patch-Based Information Reconstruction of Cloud-Contaminated Multitemporal Images”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52, No. 1, pp. 163–174.
- [6] Junchang Ju and David P. Roy, (2008) “The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally”, *Remote Sensing of Environment*, Vol. 112, No. 3, pp. 1196–1211.
- [7] Wenhui Du, Zhihao Qin, Jinlong Fan, Maofang Gao, Fei Wang, and Bilawal Abbasi, (2019) “An Efficient Approach to Remove Thick Cloud in VNIR Bands of Multi-Temporal Remote Sensing Images”, *Remote Sensing*, Vol. 11, No. 11, pp.1284.
- [8] Qiong Lu and Genyuan Zhang, (2018) “Review of Image Inpainting”, in *2018 8th International Conference on Manufacturing Science and Engineering (ICMSE 2018)*. Atlantis Press.
- [9] Zhiwei Li, Huanfeng Shen, Qing Cheng, Wei Li, and Liangpei Zhang, (2019) “Thick Cloud Removal in High-Resolution Satellite Images Using Stepwise Radiometric Adjustment and Residual Correction”, *Remote Sensing*, Vol. 11, No. 16, pp. 1925.
- [10] B. Chen, B. Huang, L. Chen, and B. Xu, (1 2017) “Spatially and Temporally Weighted Regression: A Novel Method to Produce Continuous Cloud-Free Landsat Imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 1, pp.27–37.
- [11] James Storey, Pasquale L. Scaramuzza, and Gail Schmidt, (2005) “LANDSAT 7 SCAN LINE CORRECTOR-OFF GAP-FILLED PRODUCT DEVELOPMENT”.
- [12] Quanjun Jiao, Wenfei Luo, Xue Liu, and Bing Zhang, (2007) “Information reconstruction in the cloud removing area based on multi-temporal CHRIS images”, in *MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications*, Yongji

- Wang, Bangjun Lei, Jing-Yu Yang, Jun Li, Chao Wang, and Liang-Pei Zhang, Eds. Vol. 6790, pp. 606–612, SPIE.
- [13] Bin WANG, Atsuo ONO, Kanako MURAMATSU, and Noboru FUJIWARA, (1999) “Automated Detection and Removal of Clouds and Their Shadows from Landsat TM Images”, *IEICE transactions on information and systems*, Vol. 82, No. 2, pp. 453–460.
- [14] Din-Chang Tseng, Hsiao-Ting Tseng, and Chun-Liang Chien, (2008) “Automatic cloud removal from multi-temporal SPOT images”, *Applied Mathematics and Computation*, Vol. 205, No. 2, pp. 584–600.
- [15] Huanfeng Shen, Jingan Wu, Qing Cheng, Mahemujiang Aihemaiti, Chengyue Zhang, and Zhiwei Li, (2019) “A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 12, No. 3, pp. 862–874.
- [16] Xiaoma Li, Yuyu Zhou, Ghassem R. Asrar, and Zhengyuan Zhu, (2018) “Creating a seamless 1km resolution daily land surface temperature dataset for urban and surrounding areas in the conterminous United States”, *Remote Sensing of Environment*, Vol. 206, pp. 84–97.
- [17] Rui Yao, Lunche Wang, Xin Huang, Liang Sun, Ruiqing Chen, Xiaojun Wu, Wei Zhang, and Zigeng Niu, (2021) “A Robust Method for Filling the Gaps in MODIS and VIIRS Land Surface Temperature Data”, *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15.
- [18] (2020) “Filling Gaps of Monthly Terra/MODIS Daytime Land Surface Temperature Using Discrete Cosine Transform Method”, *Remote Sensing*, Vol. 12, No. 3.
- [19] Shuang Li, Zhongqiu Sun, Yafei Wang, and Yuxia Wang, (08 2021) “Understanding Urban Growth in Beijing-Tianjin-Hebei Region over the Past 100 Years Using Old Maps and Landsat Data”, *Remote Sensing*, Vol. 13, pp.3264.

A METHOD TO COMPACTLY STORE SCRAMBLED DATA ALONGSIDE STANDARD UNSCRAMBLED DISC IMAGES OF CD-ROMS

Jacob Hauenstein

Computer Science Department, The University of Alabama in Huntsville, Huntsville, Alabama, USA

ABSTRACT

When archiving and preserving CD-ROM discs, data sectors are often read in a so-called “scrambled mode” before being unscrambled and further processed into a standard disc image. Processing of scrambled data into a standard disc image is potentially lossy, but standard disc images exhibit greater software compatibility and usability compared to scrambled data. Consequently, for preservation purposes, it is often advantageous to store both the scrambled data and the corresponding standard disc image, resulting in high storage demands. Here, a method that enables compact storage of scrambled data alongside the corresponding (unscrambled) standard CD-ROM disc image is introduced. The method produces a compact representation of the scrambled data that is derived from the standard disc image. The method allows for storage of the standard unscrambled disc image in unmodified form, easy reconstruction of the scrambled data, and a substantial space savings compared standard data compression techniques.

KEYWORDS

compact disc, compression, data archival, data preservation, scrambled

1. INTRODUCTION

In recent years, there has been increased interest in archiving and preserving software and other data produced during the previous years of computing, with especially strong interest in the archival and preservation of video games data [1–4]. Much of the work to archive and preserve such data has historically been accomplished through community efforts [2], in which a community of users work to first extract data from aging storage media (a process called *dumping* or *imaging*) and then preserve the extracted data by storing copies of it on modern storage media. The resulting data is often called a *dump*, *image*, or *disc image* [5]. Because the goal of these archival and preservation projects is to preserve the dumps over a long period of time, such dumps are typically stored on multiple media and/or in multiple locations as required for long term data storage. As such, the data storage requirements for preservation communities may grow very large, especially when dumping large media such as compact disc read-only memory discs (often denoted CD-ROMs or simply CDs), the preservation of which is the focus of this work.

Dumps are typically stored in a standard format, with the specific standard used decided upon by the community. Usage of a standard format guarantees that all community member’s dumps are in the same format, ensuring high software compatibility for each dump and enabling easy comparison between dumps from different community members via standard file hashing algorithms. In some cases, the process of dumping a medium produces two sets of data: the final dump in the standard format, and the intermediate data that is processed into the final dump. Unlike the final dump, the intermediate data is often in a format that has relatively narrow software compatibility and may be difficult to compare between community members. However, both the final dump and the intermediate data have potential importance in preservation. While the final dump is important

because it allows easy comparison of dumps between community members and has wide software compatibility (e.g., with emulators and disc image processing software that enables exploration and study of the data), the intermediate data is important because it may contain data that, due to limitations of the standard used for final dumps, is not included in the final dump. E.g., in Section 2.3, we describe in detail how data may be lost when CD-ROM dumps are processed from the often-used intermediate *scrambled* data into the standard *unscrambled* disc image used for final dumps. Thus, for the case of CD-ROM dumps, there is a need for community members to store both intermediate data and final dumps, imposing even greater storage requirements on top of the already demanding storage requirements of CD-ROM archival and preservation.

The primary contributions of this work are (1) a novel method for compactly storing the intermediate scrambled data alongside the final dump when archiving and preserving CD-ROM discs, and (2) a study of the space savings afforded by our method compared to naively storing the intermediate scrambled data. Our method can help ease the storage requirements of the CD-ROM preservation community. Our method works by attempting to reconstruct the intermediate scrambled data from the unscrambled final dump and then creating a binary diff between the reconstructed intermediate data and the original intermediate data produced during the dump.

The remainder of this work is organized as follows. Section 2 provides details about data storage on CD-ROM and the file format used for CD-ROM disc images are presented, including details about scrambling and why some data may not be preserved when the final dump is built from the scrambled intermediate data. Section 2 also presents some details about how scrambled data is dumped from CD-ROM discs, and why it is valuable to do so. Section 3 describes our method for compactly preserving the scrambled data alongside the unscrambled final dump. Section 4 describes the experiments performed to analyze the space savings of our method and the results of those experiments. Section 5 concludes the work.

2. BACKGROUND

This section presents some background details about how data is stored on CD-ROMs, why and how such data is scrambled, why there is value in dumping / preserving the scrambled data, and why it may be the case that there is data present in the (intermediate) scrambled data that is removed when a dump is converted from its scrambled form into a standard (unscrambled) image file (i.e., the final dump).

2.1. Data storage on CDs / disc images

In this section, we present some necessary background details about how information is stored on CDs and in CD disc images / dumps. Note that, because the CD specifications are quite lengthy and complex (e.g., as partially seen in [6]), we present here only enough details to aid understanding of this work. Additionally, our focus here is on the way that CDs are presented at the software level when such discs are read by standard, widely available computer optical disc drives (such as those used for dumping CDs). CDs also contain a large amount of other data (e.g., [7], [8]) at the physical layer that is not exposed / accessible at the software level by such drives and is thus outside the scope of this work and not discussed here. Since our focus here is on archival and preservation, we also assume that discs will be read in a mode that returns the most data possible from the disc. There exist reading modes that discard some error detection and correction data when reading from discs [9], but we assume those reading modes are not being used here.

Compact discs are divided into *sectors*, and each sector contains 2352 bytes. (N.B., there are some additional bytes present in the so-called *subchannels*, but we do not make use of these subchannels in this work.) When the contents of a CD are dumped and stored in a standard disc image, the disc image simply contains the 2352 bytes of every sector contained on the CD (starting with the first

sector). Thus, the logical format of CD sectors presented in this section also applies to CD disc images. (N.B., CD disc images often contain, in addition to the file that holds the sector data, other files that are used to store metadata, but we do not make use of these other files / data in this work.)

Originally, compact discs were designed for storage of stereo audio at a 16-bit sampling resolution and a 44.1Khz sampling rate, and thus 75 sectors represents 1 second of audio [9]. When storing data within a sector, the sector is divided into a number of fields. The first 12 bytes (bytes 0 through 11) of the sector are used to store the *sync field* value of 00 FF 00 hexadecimal. The remaining fields store the sector address and mode (collectively called the *header field*), user data, error correction (ECC) and detection (EDC) data, and other items.

The presence of a regular bit pattern (i.e., many more bits with a value of zero than one, or vice versa) on the physical disc is problematic for the CD decoding hardware within the optical disc drive [6]. Because such sequences may naturally occur in data, before each data sector is stored on the disc surface, the data sector is subjected to a process known as *scrambling*. In the scrambling process, each byte within the sector, except the 12 bytes that comprise the sync field, is XORed with the corresponding byte in a standardized scrambling table. The byte values contained in the scrambling table are designed to, when XORed with the sector data, avoid problematic bit patterns. The algorithm used to generate the scrambling table is standardized and described in various standards documents (e.g., [6]). This scrambling process is reversible by simply performing the same XOR a second time, and thus data is easily scrambled before the sector is written to the disc (to avoid the problematic bit patterns) and unscrambled when the sector is read from the disc (to return the data to its original state).

2.2. Reading scrambled data

When a sector is read from a CD using a standard optical drive, a sequence of 2352 bytes is returned by the drive. If the optical drive is instructed to read the sector in data mode, the optical drive (typically) automatically (1) unscrambles the data and (2) performs error detection and correction using any EDC / ECC bytes present within the data sector. Finally, the drive returns a sequence of 2352 bytes representing the unscrambled, error-corrected sector starting at the 12 byte sync field. In contrast, if the drive is instructed to read the sector in audio mode, the drive does not attempt to perform unscrambling or use any EDC / ECC bytes within the sector prior to returning a sequence of 2352 bytes representing the audio sector. In the case of reading in audio mode, the sequence of bytes returned by the optical drive typically does not begin exactly at the start of the sector. Instead, the data returned by the drive is offset by some number of bytes from the true start of the sector, with the offset amount depending on the specific optical drive model used. This audio offset has been studied widely within the optical disc archival community (e.g., [10], [11]). In addition to the audio offset exhibited by the specific optical disc drive used, some otherwise identical discs exhibit different audio offsets due to variations in manufacturing (often called the *factory offset* or *write offset*) [10]. These offsets complicate the process of archiving discs and comparing dumps between different community members / different copies of a disc, especially for discs containing both data and audio sectors (where it may be necessary to manually correct the difference in offsets between the two types of sectors [12]).

In general, optical disc drives refuse to read audio sectors in data mode (or vice versa), as this is the behavior required by the standard optical drive reading commands [9]. However, some optical drives are able to read both data and audio sectors in audio mode [13], bypassing the drive's data sector processing logic. This ability to read data sectors in audio mode (sometimes called *scrambled mode* [14]) is useful for multiple reasons. First, it ensures that the drive returns both audio and data sectors using the same offset, obviating the need for users to manually correct the offset difference between audio and data sectors. Second, it bypasses the optical drive's data unscrambling

form) in the intermediate data, but it is destroyed when the intermediate data is processed into the final dump, as shown in Fig. 1. Because as much data as possible should be preserved for archival purposes, it is thus necessary for community members to preserve both the intermediate data and the final dump, and, because the intermediate data and the final dump are each equal to the total size of the disc being dumped (i.e., they both contain all the sectors on the disc), this requirement essentially doubles the data storage requirements for each CD-ROM disc dumped (compared to storing only the final dump).

3. METHOD

In this section, we introduce our method to compactly store the intermediate scrambled data alongside the final dump when preserving CD-ROM discs. To ensure that the convenience of the final dump is not lessened when our method is applied, our method leaves the final dump unmodified and converts the intermediate scrambled data to a more compact form. This compact form can easily be used to fully reconstruct the original intermediate data.

Our method takes advantage of the fact that, for data sectors in which no EDC/ECC errors are present, the unscrambling process is exactly reversible, and such sectors can be rescrambled from the final dump into a byte sequence identical to the corresponding sector in the intermediate data. In contrast, for sectors in which EDC/ECC errors are present, the sectors are replaced with dummy sectors in the final dump (as noted in Section 2.3), and, consequently, the intermediate data for these sectors cannot be reconstructed by rescrambling the final dump. Thus, to preserve the intermediate data alongside the final dump, our method's compact representation of the intermediate scrambled data stores the intermediate data only for those sectors that cannot be exactly reconstructed from the final dump (i.e., sectors with EDC/ECC errors). Because it is typically the case that the vast majority of data sectors on a CD-ROM do not have any EDC/ECC errors, our method assumes that most data sectors can be exactly reconstructed into their intermediate format. (Our method also works for discs with a large number of EDC/ECC errors, though the space savings will be reduced.)

In the following two subsections, we describe how our method creates the compact representation of the intermediate data and how our method recreates the intermediate data from this compact representation, respectively.

3.1. Creating the compact representation

To create a compact representation of the intermediate data from the final dump, our method works in two phases. The first phase creates an approximate reconstruction of the intermediate scrambled data from the final dump. For convenience, we use ϵ to denote the file containing the original intermediate data produced during the dump and $\hat{\epsilon}$ to denote the file containing the approximately reconstructed intermediate data. For this phase, the input to our method is the disc image file containing the final dump, denoted ω , and the output is $\hat{\epsilon}$. This first phase works as follows. For each sector in ω , the sector is first checked to see if the first 12 bytes of the sector contain the sync field value. If the sync field value is not present, the sector is assumed to be an audio sector, and the sector is copied unmodified into $\hat{\epsilon}$. If the sync field value is present, each byte in the sector (excluding the 12 bytes in the sync field) is XORed with the corresponding byte in a table of the 2340 scrambling values (denoted \mathbf{T}) and then written into $\hat{\epsilon}$. (Note that, because the first 12 bytes of the sector are not scrambled, the 13th byte of the sector is the first byte that is scrambled, and it is scrambled by XORing with the 1st byte of \mathbf{T} .) This scrambling table is generated from the algorithm given in [6]. This process is performed for each sector present in ω . Upon conclusion of the first phase, $\hat{\epsilon}$ contains an approximate reconstruction on the intermediate data.

The second phase uses the *xdelta3* binary diff software [19] to encode the differences between ϵ and $\hat{\epsilon}$ into a new diff file, denoted Δ . Because, in phase 1, most sectors are exactly reconstructed from

Algorithm 1: Creating Δ from ω via $\hat{\epsilon}$

Data: \mathbf{T} **Input:** ϵ, ω **Output:** $\hat{\epsilon}, \Delta$

```

foreach 2352 byte sector  $s$  in  $\omega$  do
  if first 12 bytes of  $s$  equal sync field value
  then
    // data sector, XOR with  $\mathbf{T}$ 
    for  $i \leftarrow 12$  to 2351 do
      // XOR byte  $i$  of  $s$  with byte  $i - 12$  of  $\mathbf{T}$ 
       $s[i] = s[i] \oplus \mathbf{T}[i - 12]$ 
    end
    // copy scrambled  $s$  into  $\hat{\epsilon}$ 
    copy  $s$  into  $\hat{\epsilon}$ 
  else
    // audio sector, just copy into  $\hat{\epsilon}$ 
    copy  $s$  into  $\hat{\epsilon}$ 
  end
end
// Now that  $\hat{\epsilon}$  is constructed, use xdelta3 to diff  $\hat{\epsilon}$  and  $\epsilon$ , giving  $\Delta$ 
 $\Delta \leftarrow$  output of “xdelta3 -e -9 -s \hat{\epsilon} \epsilon”

```

the final dump into their intermediate form, ϵ and $\hat{\epsilon}$ typically differ in relatively few byte positions (as few as 0 byte positions may differ), and, as a result, Δ is typically substantially smaller than ϵ . And, because the output of *xdelta3* is Δ , a binary diff file that can be used to reconstruct ϵ from $\hat{\epsilon}$, and, because $\hat{\epsilon}$ can be reconstructed from ω , just the binary diff file Δ is sufficient to reconstruct ϵ from ω . Thus, ϵ can be discarded and the smaller Δ kept instead. Note that this approach to encoding ϵ from $\hat{\epsilon}$ is robust, because, even if some of the sectors were processed incorrectly when $\hat{\epsilon}$ was created in phase 1, Δ still contains the necessary information to rebuild ϵ from $\hat{\epsilon}$. That is, as long as phase 1 results in a $\hat{\epsilon}$ that *approximately* reconstructs ϵ at most byte positions, Δ will be smaller than ϵ . (We study the amount of space savings achieved by our method in Section 4.)

The pseudocode for both these phases is shown in Fig. 1.

3.2. Recreating the intermediate data from the compact representation

To reconstruct the intermediate data from the final dump, our method again works in two phases. The first constructs $\hat{\epsilon}$ from ω , and this phase is identical to the first phase described in the previous section. The second phase uses *xdelta3* to reconstruct ϵ using Δ .

The pseudocode for both these phases is shown in Fig. 2.

4. EXPERIMENTS AND RESULTS

In this section, we describe our experiments to evaluate the space savings of our method (compared to naively storing the intermediate data, with and without compression, alongside the final dump) and the results of those experiments.

4.1. Experiments

To study the space savings of our method, we first selected and dumped four CD-ROM discs using DiscImageCreator. As previously mentioned, DiscImageCreator dumps using scrambled mode,

Algorithm 2: Recreating ϵ from ω via $\hat{\epsilon}$ using Δ

Data: \mathbf{T} **Input:** Δ, ω **Output:** $\hat{\epsilon}, \epsilon$

```

foreach 2352 byte sector  $s$  in  $\omega$  do
  if first 12 bytes of  $s$  equal sync field value
  then
    // data sector, XOR with  $\mathbf{T}$ 
    for  $i \leftarrow 12$  to 2351 do
      // XOR byte  $i$  of  $s$  with byte  $i - 12$  of  $\mathbf{T}$ 
       $s[i] = s[i] \oplus \mathbf{T}[i - 12]$ 
    end
    // copy scrambled  $s$  into  $\hat{\epsilon}$ 
    copy  $s$  into  $\hat{\epsilon}$ 
  else
    // audio sector, just copy into  $\hat{\epsilon}$ 
    copy  $s$  into  $\hat{\epsilon}$ 
  end
end
// Now that  $\hat{\epsilon}$  is constructed, use xdelta3 to apply  $\Delta$  to  $\hat{\epsilon}$ , giving  $\epsilon$ 
 $\epsilon \leftarrow$  output of “xdelta3 -d -s  $\hat{\epsilon} \Delta$ ”

```

producing both scrambled intermediate data and a final unscrambled dump. We then, for each of the four discs, used our method to generate the compact representation (i.e., Δ) of the intermediate scrambled data. Finally, we compared the size of the compact representation generated by our method with the size of the original scrambled data from the corresponding final dump. In addition, we compared the size of our compact representation with the size of the intermediate data when it is compressed using 7-Zip’s “Ultra” mode [20]. To compare sizes, the *space saving*, denoted k , was calculated from the new size (i.e., the size of the 7-Zip compressed file or Δ) and the original size (i.e., the size of the original scrambled data) according to

$$k = 1 - \frac{\text{New Size}}{\text{Original Size}}. \quad (1)$$

Thus, k is equal to 0 if the new size and original size are equal (i.e., when there is no space savings) and increases as the amount of space savings goes up.

The four discs dumped, denoted **D1**, **D2**, **D3**, and **D4**, were selected such that they represent a variety of possible disc types that may be input to our method. **D1** contains data sectors only and does not contain any intentional error sectors. **D2** contains both data sectors and audio sectors and does not contain any intentional error sectors. **D3** and **D4** both contain data sectors only, and both contain intentional error sectors.

4.2. Results

The results are summarized in Table 1. There, the original size and number of error sectors are shown for each disc. In addition, the size of the 7-Zip compressed scrambled data is shown and the size of Δ when the scrambled data is compacted according to our method is shown. Finally, the space savings value k is shown for both 7-Zip and our method.

As can be seen in the table, our method achieves a much higher space savings value compared to 7-Zip. As seen here, because the scrambling process helps to ensure that data has similar numbers

Disc	Original Size	No. Err. Sec.	7-Zip Size	Our Method Size	7-Zip k	Our Method k
D1	493,146,192	0	459,200,084	1,718	0.069	0.999
D2	752,425,968	0	580,962,585	2,612	0.228	0.999
D3	788,886,672	585	661,481,962	1,261,927	0.161	0.998
D4	830,822,832	583	828,265,184	1,297,480	0.003	0.998

Table 1: Results of our method and 7-Zip Ultra compression on the four discs. All sizes in bytes.

of bits with values of 1 and 0, the scrambled intermediate data has a high level of entropy and typically does not compress well via standard compression algorithms. In contrast, our method exhibits a very high space savings.

5. CONCLUSION

In this work, we introduced a new method for compactly storing intermediate scrambled data alongside final dumps. Our method takes advantage of the fact that the intermediate scrambled data can be approximately reconstructed from the final dump. Our method first builds an approximate reconstruction of the scrambled intermediate data from the final dump, and then encodes the differences between the approximate reconstruction and the actual intermediate data.

Our method achieved a substantial space savings increase compared to storing the intermediate data without compression and compared to storing the intermediate data using 7-Zip’s Ultra compression. Thus, we believe our method will prove useful for easing the data storage burden encountered by those archiving and preserving CD-ROMs.

ACKNOWLEDGMENTS

We wish to thank the members of the data archival and preservation communities. We also wish to thank the reviewers for their helpful feedback.

REFERENCES

- [1] M. Guttenbrunner, C. Becker, and A. Rauber, “Keeping the game alive: Evaluating strategies for the preservation of console video games,” *International Journal of Digital Curation*, vol. 5, no. 1, Jun. 2010. [Online]. Available: <https://doi.org/10.2218/ijdc.v5i1.144>
- [2] F. Cifaldi, ““It’s just emulation!” - The challenge of selling old games,” in *Game Developers Conference*, 2016. [Online]. Available: <https://www.gdcvault.com/play/1023470/contactUs>
- [3] J. Newman, “The music of microswitches: Preserving videogame sound—a proposal,” *The Computer Games Journal*, vol. 7, no. 4, pp. 261–278, 2018. [Online]. Available: <https://doi.org/10.1007/s40869-018-0065-8>
- [4] N. Nylund, P. Prax, and O. Sotamaa, “Rethinking game heritage—towards reflexivity in game preservation,” *International Journal of Heritage Studies*, vol. 27, no. 3, pp. 268–280, 2021. [Online]. Available: <https://doi.org/10.1080/13527258.2020.1752772>
- [5] Redump.org Community, “Redump.org,” 2022, accessed Jun. 15, 2022. [Online]. Available: <http://wiki.redump.org/index.php?title=Redump.org&oldid=48927>
- [6] “Data interchange on read-only 120 mm optical data disks (CD-ROM),” Ecma International, Geneva, Switzerland, Standard, Jun. 1996. [Online]. Available: <https://www.ecma-international.org/publications-and-standards/standards/ecma-130/>
- [7] L. B. Vries and K. Odaka, “CIRC—the error-correcting code for the compact disc digital audio

- system,” in *Audio Engineering Society Conference: 1st International Conference: Digital Audio*. Audio Engineering Society, 1982.
- [8] K. Immink, “Modulation systems for digital audio discs with optical readout,” in *ICASSP '81. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, 1981, pp. 587–589.
- [9] “Multimedia command set - 5 (MMC-5),” T10, Washington, D.C., USA, Standard, Oct. 2006. [Online]. Available: <http://www.t10.org/cgi-bin/ac.pl?t=f&f=mmc5r04.pdf>
- [10] Redump.org Community, “Combined offset in eac,” 2010, accessed Jun. 15, 2022. [Online]. Available: <http://forum.redump.org/topic/7649/combined-offset-in-eac/>
- [11] Accuraterip.com, “Accuraterip,” 2010, accessed Jun. 15, 2022. [Online]. Available: <http://www accuraterip.com/>
- [12] Redump.org Community, “CD dumping guide with audio tracks (old),” 2022, accessed Jun. 15, 2022. [Online]. Available: [http://wiki.redump.org/index.php?title=CD_Dumping_Guide_with_Audio_Tracks_\(Old\)&oldid=46420](http://wiki.redump.org/index.php?title=CD_Dumping_Guide_with_Audio_Tracks_(Old)&oldid=46420)
- [13] —, “DiscImageCreator: Optical disc drive compatibility,” 2022, accessed Jun. 15, 2022. [Online]. Available: http://wiki.redump.org/index.php?title=DiscImageCreator:_Optical_Disc_Drive_Compatibility&oldid=48878
- [14] sarami, “DiscImageCreator,” <https://github.com/saramibreak/DiscImageCreator>, 2022, accessed Jun. 15, 2022.
- [15] K. Kaspersky, *CD Cracking Uncovered: Protection Against Unsanctioned CD Copying*. Wayne, PA, USA: A-List Publishing, 2004.
- [16] Human Head Studios, “Rune,” 2000, accessed Jun. 15, 2022. [Online]. Available: <https://web.archive.org/web/20130622083143/http://www.rune-world.com/>
- [17] Redump.org Community, “Moderating guidelines for IBM PC and other systems,” 2021, accessed Jun. 15, 2022. [Online]. Available: http://wiki.redump.org/index.php?title=Mode rating_guidelines_for_IBM_PC_and_other_systems&oldid=45839
- [18] —, “Issues dumping pc disc with “code lock” copy protection,” 2021, accessed Jun. 15, 2022. [Online]. Available: <http://forum.redump.org/topic/29842/issues-dumping-pc-disc-with-code-lock-copy-protection/page/2/>
- [19] J. P. MacDonald, “xdelta: open-source binary diff, differential compression tools, VCDIFF (RFC 3284) delta compression,” 2016, accessed Jun. 15, 2022. [Online]. Available: <http://xdelta.org/>
- [20] I. Pavlov, “7-Zip,” 2021, accessed Jun. 15, 2022. [Online]. Available: <https://www.7-zip.org/>

AN OVERVIEW OF PHISHING VICTIMIZATION: HUMAN FACTORS, TRAINING AND THE ROLE OF EMOTIONS

Mousa Jari^{1, 2}

¹School of Computing, Newcastle University, Newcastle, UK

²College of Applied Computer Science,
King Saud University, Riyadh, Saudi Arabia

ABSTRACT

Phishing is a form of cybercrime and a threat that allows criminals ('phishers') to deceive end-users in order to steal their confidential and sensitive information. Attackers usually attempt to manipulate the psychology and emotions of victims. The increasing threat of phishing has made its study worthwhile and much research has been conducted into the issue. This paper explores the emotional factors that have been reported in previous studies to be significant in phishing victimization. In addition, we compare what security organizations and researchers have highlighted in terms of phishing types and categories as well as training in tackling the problem, in a literature review which takes into account all major credible and published sources.

KEYWORDS

Phishing, emotion, information, victimization, training.

1. INTRODUCTION

Phishing is a kind of social engineering attack that is used to steal an individual's data, including personal identification details, credit card numbers or any other credentials. This activity occurs when a phisher pretends to be someone who is a trusted individual and persuades a victim to open a certain email or a message. When a victim opens such a communication, his information can be hacked/leaked and made available to the email/message sender. McAlanay and Hills described phishing as a social engineering tool or a threat that causes a risk to cyber security [3]. They further highlighted that phishing emails or messages are based on the assertion of some urgency or threat where an attacker or phisher causes a victim to become blackmailed having been encouraged to respond to the email or message accordingly [3]. According to Shaikh et al., phishing is a serious threat in the cyber world that is causing billions of dollars of losses to internet users through the use of social engineering and technology by gaining access to their financial information [4]. In phishing, the attacker sends spoof emails to the internet user which deceives victims and causes them to disclose their sensitive and confidential data [4]. Consequently, from an analysis of the relevant research based on the common elements which have been identified, one can define phishing as a cyber threat which, due to the deployment of social engineering techniques and technological means, leads to messages and emails being sent to internet users resulting in the retrieval of personal information about victims, hence causing them monetary or other damage through the leaking of information.

In this research, the aim is to identify human factors, and specifically emotional variables, which lead to a higher probability of phishing victimization. This problem has been discussed in various studies, and so the method employed in this paper is to analyse the literature review and secondary research in order to highlight the emotional factors which play a significant role in phishing victimization, as well as comparing how security organizations define and address phishing and provide advice on how to avoid becoming a victim.

2. OVERVIEW OF PHISHING

Phishing is a relatively new concept which was first employed in the late 1990s, and in the past years there has been an increasing trend of damage caused by phishing. Rather than simply the deployment of technical expertise to attempt to successfully compromise system security, phishing can also be defined as a semantic attack that uses social engineering tactics to persuade internet users to disclose their private and confidential information such as login credentials, social security numbers, and bank account details. Phishers most commonly use an e-mail which includes an embedded hyperlink with a message either sharing some threat, such as a warning message about account closure, or reporting positive news, for example hinting at an unclaimed reward, to attract a potential victim. When a person clicks on the malicious link, it leads to a web-based form that mimics those of valid and authentic websites asking a user to enter login credentials. Once added, such information is then used to compromise network security and thus sensitive information reaches the phisher. When phishers have retrieved sensitive information from victims, they can then either sell the information, open bank accounts, or even steal the victim's money. Such phishing attempts are believed to be the 'vector of choice' among cybercriminals.

The Anti-Phishing Workgroup has discovered up to 40,000 phishing websites per month, targeting almost about 500 unique brands; however, the US Department of Defense and the Pentagon have reported more than 10 million phishing attacks each day, which is clearly a huge number [4][6]. However, Harrison also highlighted the fact that the success rate of phishing attacks is never 100% and often varies between 30-60% [4][6]. This research study aims to explore what makes these attempts successful, focusing on the emotional factors that may lead users to share confidential information with someone unknown to them.

2.1. Types of Phishing According to Research and Security Organizations

Researchers and the security sector have listed various types of phishing through which the phishers target internet users. In a paper entitled 'Fifteen years of phishing: can technology save us?' Furnell et al. highlighted four major types: spear phishing, clone phishing, whaling, and bulk phishing [3][5]. In spear-phishing, specific individuals or companies are targeted using a tailored message. In this type of scam, the attacker is more likely to have some background information about the target, based on which the message that is created becomes more convincing and successful in deceiving the recipient. Therefore, users become more likely to be targeted and lose their confidential information. Meanwhile, clone-phishing attackers make use of a valid email that contains a URL or attachment and retain the content of the actual message. However, the embedded link or attachment is replaced with a malicious file so that the original sender is spoofed due, for example, to a claim that the email message is an update of an earlier version. In contrast, whaling is considered to be a particularly threatening type of phishing where CEOs or other senior or high-value individuals in an organization are targeted. Here, the medium used in communicating the message is still the email, but in addition similar kinds of threats are also sent through 'vishing' (voice phishing) and 'smishing' (SMS phishing) which are terms used to specify threats via voice telephony and text messaging. The final category of bulk phishing

occurs when there is no specific target or any tailored message. The approach employed instead is to send bulk emails to as many users as possible, and the success of this kind of scam depends on such large-scale mailing where a sufficiently large number of recipients mistakenly believe that the email is relevant to them [3][5].

Various security organizations have categorized phishing in different ways, but the forms listed overlap with the four major types indicated above. Reports published by the US Federal Trade Commission, the Surveillance Self Defense and Get Safe Online groups, and Phishing.org specify 14 major types of phishing, including: spear phishing, session hijacking, email spam, content injection, web-based delivery, phishing through a search engine, link manipulation, vishing, smishing, key logging, malware, trojans, ransomware, and malvertising [5–10]. The common element in all phishing categorizations provided by security organizations and Furnell et al.'s research is the fact that attackers use email, voicemail, or SMS to accomplish phishing [5]. Similarly, victims in all types vary from ordinary internet users to specific companies or high-profile individuals. In addition, the concepts exploited in all the phishing types overlaps with the categories recognized by the security organizations, except for web-based delivery, phishing through a search engine and key logging. In web-based delivery phishing, which is also called 'man-in-the-middle' phishing, the attacker is positioned between the customer and the original website. Using his phishing system or network, the phisher identifies the confidential information of the victim during the completion of a deal between the user and the legitimate website. In the case of search engine-based phishing, the user is captured by being (re-)directed to websites purportedly offering low-priced products or services. When the user tries to purchase a product by adding his credit card details, the data are collected by the phisher. Moreover, key logging phishers identify keyboard strikes and mouse clicks performed by a user, and from this information they manage to retrieve passwords and other confidential data[3][5].

2.2. Training and Increasing Awareness

Online safety and the avoidance of all threats that remain present around users is not easy. In the case of phishing, this is a specific kind of scam that plays with human psychology and attracts the attention of victims using various techniques to cause the damage explained above. Questions therefore arise concerning how people can be trained to stay safe from phishing, and publications from various researchers and security organizations addressing this issue are compared here in order to draw conclusions about how to protect users against the rising threat of phishing.

Jensen et al. considered aspects of training which might mitigate the impact of phishing attacks, focusing on the use of 'mindfulness' techniques [9][11]. The authors specified that simple decision making or mental shortcut methods to avoid phishing are no longer effective, since not only are they short-term strategies but also phishers are now very familiar with such models. So, the researchers wanted to create an innovative approach to training which teaches internet users to develop new mental models and strategies for the allocation of attention when examining online messages. Rather than a rule-based approach which repeats multiple rules and cues, the training was designed as an exercise to enhance the degree to which users attend to and understand the approach being used in the received message; in other words, to promote 'mindfulness' in evaluating the message. The concept of mindfulness here concerns paying receptive attention to one's experience and surroundings so as to improve the ability to understand one's environment along with an enhanced self-regulation capacity and stronger behavioural control. Their training module consisted of graphics in addition to text promoting mindfulness, and helped provide a better understanding of how to avoid phishing attacks, since the graphical representation of concepts is thought to enhance the capacity to acquire information leading to better performance in complicated tasks [9][11]. The project aimed to provide participants with a blend of mindfulness training techniques and a rule-based approach so that

they could respond to phishing attacks more effectively. To test the effectiveness of the training, a dummy phishing attack was launched in which the participants were directed towards a fictitious website where they were asked to enter their university account login credentials. The results indicated that the graphic and text-based training formats was equally successful in decreasing participants' probability of responding to phishing messages in such a way as to become victims. However, the approach using mindfulness significantly reduced the likelihood of participants responding to the phishing messages, and hence was found to be useful against phishing [11].

Wash and Cooper have also explored the training models that can work best against phishing [12]. The researchers indicate that raising awareness among users through facts-and-advice training or storytelling models works better in combating phishing, but using professional security experts or peers to deliver such training is needed to make it more effective. The methodology of the study involved sending 2000 participants phishing emails to gather data necessary to assess the effectiveness of the activity. The lessons that could be learned included to "type in URLs; don't click on them" and "look for HTTPS", that "misspellings can signal fake emails" and "phishing is your problem; don't rely on others to protect you". In total, 17 lessons were compiled from the results of the activities performed. It was found that not all of the lessons helped in combatting phishing attacks, but all were considered significant enough to be presented to the participants when using the fact-telling approach. Meanwhile, the comparison with a story-based strategy led to the surprising outcome that the facts-and-advice approach resulted in fewer clicks leading to scams when an expert was used for the training, whereas the storytelling approach also resulted in lower click rates but only when peers were used rather than experts [12]. From the above-mentioned studies it can be concluded that multiple factors may lead a user to become caught by a phishing attack, and that appropriate training and education is needed in order to be safer.

With respect to the benefits of such education, the conclusions drawn in a study by Chaudhary include a number of significant recommendations as follows [13]:

- I. Providing any new knowledge in order to be up to date is important, but security education should also result in eliminating misconceptions relating to security.
- II. The security education that should be part of a curriculum needs to be up-to-date, and it should cover both new technologies, and information about sophisticated phishing threats and attacks.
- III. Security education must impart knowledge related to technological and non-technological attacks and threats.
- IV. The design of curricula for security education should be based on the input of relevant stakeholders, including teachers, learners, and IT and security professionals, since their experience, skills and knowledge can help in covering a wide range of security-related topics.
- V. The adoption of a more interactive way of teaching and learning methods can be quite helpful in making both security learning and teaching more interesting and potentially effective [13].

In parallel with academic researchers, many security organizations have also put a lot of emphasis on user security against phishing attacks, because its severity can vary from password retrieval to stealing money, ultimately causing considerable damage. Among the most prominent security organizations is the National Cyber Security Centre (NCSC), which is a UK-based organization that provides support to critical organizations, including many in the public sector and industry as well as the general public.

In response to the rapid increases in cyber threat levels, the NCSC provides efficient incident responses to mitigate harm and facilitate recovery, and compiles information on the lessons learned that can be useful in the future. Apart from providing solutions to possible phishing threats, NSCS is also concerned with educating people to develop self-reliance against phishing attacks. For this purpose, a major contribution of the Centre is the design of practical resources for school students who take an interest in cyber security studies. The projects on which NCSC is working to provide cyber security education include the CyberFirst courses, schools, colleges, bursaries and apprenticeships and associated resources, and the CyberSprinters programme [14].

Similarly, the Get Safe Online organization provides users with a greater awareness of phishing and online scams through its informative online articles and shares tips and tricks to raise consciousness among users concerning phishing attacks. Advice is given, for example, on how to use emails wisely, how to identify fraudulent emails, how to distinguish between legitimate and phishing websites and emails and, if one has lost money due to an online scam, what course of action to take [10].

The Surveillance Self-Defense organization is also based on providing general public protection from phishing attacks. Its literature specifies the intensity of malware and its role in introducing phishing threats. The implementation of malware by phishers is usually based on stealing passwords, where the malware is installed when a user opens or clicks on a malicious link, downloads an unknown file, visits a compromised websites, downloads automatic content, or even when USBs are shared while plugging into suspicious ports. However, despite the multiple ways through which malware can be used for phishing [9], users can be educated to avoid being a phishing victim by implementing five important measures:

1. Updating systems and using licensed software.
2. Backing up data.
3. Pausing before clicking, and thus to be more vigilant and to avoid clicking immediately.
4. Using full-disk encryption along with a strong password.
5. Using better anti-virus techniques.

From the discussion above and in the light of the relevant research, it can be concluded that the frequency and intensity of online scam, phishing, and fraud activity are increasing with the passage of time, and so, in order to be safe, security education and training are necessary and perhaps should be mandatory.

2.3. The Role of Emotion

Chaudhary has emphasized the role of emotion in his research [13], explaining that the manipulation of emotion is generally found to be a prime target of phishers. Ignorance, a desire to be liked, gullibility, and wanting to be helpful to others are among the aspects associated with emotionality which are commonly targeted by scammers or phishers, who rely on the exploitation of vulnerability and weakness. People are found to be more inclined towards sharing their information with others when strong emotions have been triggered, and human behaviour when triggered this way is more likely to be driven by subconscious processes. The problem here is that the functionality of the subconscious mind is not based on logical or analytical behaviour, a fact which is exploited by phishers in pursuing their aims [13]. However, although emotions are very important and can be used against a victim as a weakness, they can also act in the victim's favour as a strength too. If the emotions which are exploited by phishers instead remain under the control of the user, this may help to combat phishing, which implies that emotions should also have a significant role in training and awareness-raising.

Chaudhary has discussed a very interesting type of phishing and social engineering attack in his research. This is called farming, where a phisher develops a relationship with a victim and continues to obtain relevant information over a certain period of time [15][13]. This activity is usually conducted in four phases. The first phase is information gathering, which involves the collection of the necessary data so that a relationship with a potential victim can be built. The second phase is based on developing the relationship, such as by coordinating with the victim and building a trust-based connection. In the third phase, exploitation starts where the victim is manipulated and deceived to obtain critical desired information, and the final phase is the execution of an attack using the information provided, to the detriment of the victim but beneficial to the attacker.

Emotion may exert a significant influence on many human cognitive processes such as attention, perception, memory, learning, reasoning and questioning, and problem-solving. If a person can manage to understand his emotions and learns how to control them, he can understand his surroundings better, communicate more efficiently, and even appreciate the worth of any relationship [16]. In relation to phishing, it can therefore be proposed that, if internet users are provided with the training and awareness based on emotional control, then they will be less likely to be successfully targeted by phishers.

To mitigate the malevolent exploitation of emotions, Jaeger and Eckhardt have highlighted the significance of emotions in awareness-raising and training [15]. They believe that human emotions are learnt, and when they are taken under appropriate control the impact of phishing attacks may be overcome. The researchers analyzed the relationships among the constructs of protection-motivation theory (PMT) Nomology [22] that involve fear and the motivation for protection and in actual security-related behaviour, indicating that perceived threat perceived coping efficacy in response to threat encourages a person's motivation towards self-protection in combatting the threat. So, when an individual encounters a phishing attack and faces its likely consequences; then, after being threatened, he starts to believe that he can respond to the situation using learnt behaviours and emotions which can ensure protection against such threats in the future [15]. Moreover, this helps not only in terms of training but also in creating awareness among peers. Such learning can also lead to technical solutions, such as users protecting themselves from phishing by implementing technical countermeasures including deciding not to click on any unknown or potentially malicious link, not downloading an .exe file, and deleting any dubious email or sending it to the junk folder.

2.4. Human Factors in Phishing Victimization

If one asks what makes phishing successful or what causes victims to become entrapped in phishing, the answer is simple: the victim himself. More specifically, it can be said that phishers target the victim's emotions which they manipulate to achieve their aims [16]. The above sections have indicated that emotions have a major role in phishing attempts, and the focus of the remaining discussion is to answer the research question of the study: what are the human factors, and specifically the emotional variables, which lead to phishing victimization.

In considering the nature of emotions and other psychological variables, Chaudhary cited various aspects of the human psyche which play a major role in phishing victimization [13] and specified several psychological states and factors that are mainly targeted by phishers which may lead a user to comply with the instructions given as part of the phishing attempt [13]. These include:

- i) Reciprocation: where potential victims are more likely to comply with malicious instructions when they have a feeling of gratitude towards the phisher and feel that they are granting a favour to one in need.

- ii) Consistency and commitment: since people like to be seen as trustworthy by fulfilling promises. If this trait is targeted by the phisher, to make one feel that he has made a promise, then it is possible that the person may comply with the phisher's instructions and demands.
- iii) Social proof: people may be deceived more easily if they are provided with persuasive evidence, such as being convinced that one is not alone in doing something and everyone else is doing the same thing, so that trapping a victim becomes more likely.
- iv) Liking: using the emotion of liking someone is often exploited as a tool by phishers, because people more readily comply with someone they like. If a phisher manages to masquerade as a person the victim likes, the phishing attempt could succeed.
- v) Authority: people generally comply with authority, since being a responsible citizen usually means complying with an authorized person. So, if a phisher manages to appear authoritative, he can use the victim's tendency to comply with the demands of an authority to manipulate him.
- vi) Scarcity: if a phisher manages to convince his target that something he wants is in short supply and will not be available afterwards, then it is more likely that the victim may comply with the phisher's instructions.

In addition to the above-mentioned psychological states and emotions described by Chaudhary, Vishwanath et al. considered the dimensions of the email and social media behaviour of individuals which result in getting trapped by phishing attempts [17]. They highlighted the fact that social media users, and especially those who regularly check Facebook notifications, are more likely to be targets of social media phishing. However, social media are quite distinct in providing relational information which can help in the detection of deception. Social media-based phishing attacks are multi-staged in the sense that the user receives a friend-request followed by messages. This is in contrast to email-based phishing which is single-staged, where a phisher uses a persuasive subject line which either causes a feeling of fear in cases of a threat, or a sense of happiness following a piece of fake news such as concerning winning a lottery or an amount of money. Vishwanath et al. concluded that users with low levels of emotional stability are more likely to start worrying and lose their emotional control based on the subject of the email. This kind of behaviour can create impulsive email habits. For example, in response to the sound of a single email notification, a user may react by checking and immediately opening an email to answer it, which may lead to reactively clicking on malicious phishing links [17]. Responding to emails with feelings of being nervous, curious, happy or under threat has also been explored by other researchers because this area of research has shown considerable promise.

Abroshan et al. recently conducted a noteworthy research study regarding human behaviour and emotions which influence the success of phishing attacks' [18]. They highlighted previous studies which have found that emotional behaviour can significantly affect responses to phishing emails, and proceeded to develop a holistic method including the use of psychological and phishing mitigation to identify highly susceptible users in organizations who are at the risk of clicking on phishing emails. Their proposed solution is comprised of three modules involving behaviour measurement, risk scoring and mitigation, and the system can be delivered online. It is also a flexible solution which can be expanded by adding more human factor root-causes; for example, "more behavioural and emotional factors that might impact falling into a phishing scam" (p. 349). This study significantly highlights the importance of human behaviour and emotions in relation to security behaviour such as the propensity to get caught up in a phishing scam. For example, the emotions of users such as fear and anxiety, especially in certain situations like the Covid-19 pandemic, can play a pivotal role in making phishing attacks successful. This is because the user's awareness and knowledge of security can be overshadowed due to the emotion of fear [18]. In reacting, users might click on a suspicious phishing link without thinking, supposing that the information is required due to Covid-19 health impacts.

2.5. User Knowledge, Education, and Understanding

Dealing with phishing attempts can be controlled through the use of software; however, the best prevention can only be provided through the user's improved knowledge, education and understanding. In one study, Arachchilage and Love emphasized that anti-phishing education and knowledge needs to be considered in order to combat phishing [19], and they investigated the extent to which procedural knowledge or conceptual knowledge has positive effect on users' self-efficacy to be safe from phishing attacks. Using a theoretical model based on Technology Threat Avoidance Theory, data was collected from 161 computer users who were provided with a questionnaire to get their feedback. It was found that both procedural and conceptual knowledge positively impacted the users' self-efficacy, ultimately resulting in the enhancement of their phishing threat avoidance behaviour. A later study by He and Zhang supported the claim that users' knowledge, education and understanding play a significant role in repulsing phishing attacks, and the authors concluded that "Training programs and educational materials need to relate cyber awareness to employees' personal life, family, and home, in order to be more engaging and to encourage employees to change their cybersecurity behaviour" [20].

Subsequent research regarding knowledge capabilities was conducted by Wash et al., who surveyed 297 participants with matching demographic characteristics in the US, allowing them to share their experiences of phishing emails [21]. This study provides evidence that humans may perceive and experience phishing emails in a very idiosyncratic manner, using different capabilities and knowledge in contrast to technical filters. For example, their past knowledge may assist them in detecting and becoming suspicious of phishing attacks, such as their familiarity with previously received emails as well as their expectations regarding incoming emails. It can be assumed that this knowledge is contextual and every individual will have a unique set of relevant experience, which will be utilised to detect missing and unexpected informational units in emails. Because technical solutions seldom spot these types of phishing attacks, such knowledge-based information residing only in the human mind is critical in spotting phishing attacks, whereas technical expert-based filters lack this information processing capability. For example, humans can use their knowledge to conduct an investigation or delay the response to an email and request further information from the sender [21]. This shows that the user's knowledge is critically important in combatting phishing attacks.

From a linguistic perspective, although it is acknowledged that those who are non-native speakers are more vulnerable to phishing attacks [23], most published studies fail to consider language-based phishing vulnerability. However, Hasegawa et al. conducted a noteworthy online survey of 302 Japanese, 276 South Koreans and 284 German participants representing a total of 862 non-native English speakers. The results of the analysis of data revealed that participants who were not familiar or confident with the English language had a high propensity to ignore all emails written in the English language [23]. Additionally, a qualitative analysis revealed five key factors that aroused the concern of participants in identifying phishing emails in English. These include difficulties in identifying errors in the language, unfamiliarity with the written English in phishing content, difficulty in understanding English content, and decreased attention. These findings suggest that it would be necessary to develop different strategies to tackle the susceptibility to phishing emails among non-native speakers, as well as to consider the importance of language barriers when formulating interventions to assist non-native speakers to combat phishing attacks.

2.6. Demographics Factors

In addition to the emotional factors discussed by Chaudhary and others, various demographic variables are believed to be significant in phishing victimization. However, other demographic

characteristics have been found to have an impact on resilience against phishing. For example, Gopavaram et al. found that phishing resilience and age have a negative relationship, so that older users are more likely to become confused about the legitimacy of genuine and phishing websites. However, no significant relationship was identified between phishing resilience and gender [24], whereas Sheng et al. [25] found that age and gender are key demographic factors that can indicate the levels of susceptibility to phishing. Their analysis indicated that women click on malicious links provided in phishing emails more often than men, and hence are more likely to provide phishers with confidential information. Such differences in gender-based behaviour might be based on the role of technical education, since males often have more technical knowledge than females. Age was also found to have a significant relationship with phishing susceptibility, and participants between the ages of 18-25 years were more likely to fall into phishing traps. But this factor was also linked with levels of education, since participants in the age group concerned were found to have received a relatively lower level of education, limited training material, fewer years spent using the Internet, and low capacity in risks management [25]. So, in relating demographics to phishing, Arachchilage and Love's conclusion that improved education can reduce phishing victimization was supported.

2.7. Online Habits and Behaviour, Responding Impulsively to Emails, and the Role of Mental Models

In research where the scope and purpose are to understand the phishing attacks and victims' reasons for opening malicious links, an investigation of the mental models used and online behaviour exhibited is very significant. A simple explanation of a mental model is an individual's own thought processes in relation to how a particular phenomenon works in a real-life scenario. Mental models are based on an individual's learning, experience, skills and knowledge which improve the thought process concerned, hence resulting in some specific behaviour or outcome. So, no single mental model can work against phishing but several mental models can serve the purpose, and Jaeger and Eckhard have explained that schemata and mental models are key elements used to achieve high levels of awareness [15]. Critical cues that are needed to activate both of those mechanisms should be based not only on the characteristics of emails but should also be linked to security warning alerts. Meanwhile, mental models specifically relating to phishing could be more complex, since they may vary depending on the type of phishing attacks concerned. An individual's mental model is shaped by past experience where phishing plays a major role in obtaining situational information relating to security awareness. It was concluded that experienced users signify their level of awareness by using the security-related information cues available, which shows that the experience helps in developing schemata and remembering the critical cues, ultimately leading to pattern matching and improved thought processes in working against phishing [15]. In another study, Sibrian et al. conceptualized the thought processes and human behaviour involved using a model of the social decision-making process divided into two systems [26]. The first was defined as the source of emotional reactions based on experience, which works quite rapidly and almost impulsively with very little voluntary control, whereas the second system is based on reasoning, focus, and choice. The operations linked to this second system require attention, and can be disrupted when it is reallocated or disturbed. Since it is more logical and rational, whereas the emotional system is more impulsive, phishers exploit it so that the other system either does not react or takes too much time to respond. This is how phishers exploit the human mental model, due to which an individual may feel fear, happiness, curiosity or even urgency to respond to the phishing message quickly, which affects the overall user's behaviour while responding to a phishing attack and leads to getting entrapped in a phishing attack [26].

2.8. Expert versus Non-expert Thought and the Ability to Detect and Avoid Falling Victim to Phishing

The behaviour of experts and non-experts to phishing attacks differ, which primarily depends on one's skills, knowledge and experience. Nthala and Wash explain that non-experts follow four sense-making processes according to which they determine if the email they receive is in actuality, a phishing message [27]. In the first stage, they identify, without going into detail, whether or not the email is relevant to them. In the second stage, the goal is to understand why they received the email. Here, non-experts try to understand the email in more depth. In the third stage, they start to take a positive action using a sense-making process to fulfil the request made in the email. In the last stage, either the email is closed, deleted, or even moved to re-reading. The core element of this stage is the sense of marking the task as completed by closing the email [27].

As opposed to non-expert behaviour, Wash highlighted that, to identify a phishing email, experts follow a three-stage process [28]. In the first stage, experts tend to consider why such an email has been received and how it is relevant to them, and identify any discrepancies. In stage two, they may entertain suspicions about the email by analysing its features such as the presence of a link requiring a click. Here, they manage to identify that the email is based on a phishing message. In the third step after this thought process using their mental model, experts deal with the phishing email either by reporting or deleting it [28].

3. CONCLUSIONS

Phishing is one of the most common problems which causes an individual to lose confidential information such passwords or credit card numbers which may even trigger the theft of money. In this paper, we compared what security organizations and researchers have emphasised in terms of phishing types and categories as well as training and awareness in tackling the problem. Numerous emotional factors are targeted by phishers; however, with the help of training and anti-phishing education, emotions can be managed and controlled, and self-control can be developed that can lead to phishing attempts being unsuccessful.

REFERENCES

- [1] G. Sonowal, "Introduction to phishing," *Phishing and Communication Channels*, Apress, Berkeley, CA, pp. 1–24, 2022, doi: 10.1007/978-1-4842-7744-7_1.
- [2] R. G. Brody and F. St Petersburg Valerie Kimball, "Phishing, pharming and identity theft," *Academy of Accounting and Financial Studies Journal*, vol. 11, no. 3, 2007.
- [3] J. McAlaney and P. J. Hills, "Understanding phishing email processing and perceived trustworthiness through eye tracking," *Frontiers in Psychology*, vol. 11, p. 1756, Jul. 2020, doi: 10.3389/FPSYG.2020.01756/BIBTEX.
- [4] A. N. Shaikh, A. M. Shabut, and M. A. Hossain, "A literature review on phishing crime, prevention and investigation of gaps," *SKIMA 2016 - 2016 10th International Conference on Software, Knowledge, Information Management and Applications*, pp. 9–15, May 2017, doi: 10.1109/SKIMA.2016.7916190.
- [5] S. Furnell, K. Millet, and M. Papadaki, "Fifteen years of phishing: can technology save us?". *Journal of Computer Fraud and Security*, vol. 2019, no.7, pp.11–16, Nov. 2021, doi:10.1016/S1361-3723(19)30074-0.
- [6] B. Harrison, E. Svetieva, and A. Vishwanath, "Individual processing of phishing emails: how attention and elaboration protect against phishing," *Online Information Review*, vol. 40, no. 2, pp. 265–281, Apr. 2016, doi: 10.1108/OIR-04-2015-0106/FULL/PDF.
- [7] Get Safe Online, "Spam and scam email: Get Safe Online." <https://www.getsafeonline.org/personal/articles/spam-and-scam-email/> (accessed Feb. 17, 2022).

- [8] “How to recognize and avoid phishing scams: FTC consumer information.” <https://www.consumer.ftc.gov/articles/how-recognize-and-avoid-phishing-scams> (accessed Feb. 17, 2022).
- [9] Surveillance Self-Defense, “How to: avoid phishing attacks.” <https://ssd.eff.org/en/module/how-avoid-phishing-attacks> (accessed Feb. 17, 2022).
- [10] Phishing.org, “Phishing: what is phishing?” <https://www.phishing.org/what-is-phishing> (accessed Feb. 17, 2022).
- [11] M. L. Jensen, M. Dinger, R. T. Wright, and J. B. Thatcher, “Training to mitigate phishing attacks using mindfulness techniques,” *Journal of Management Information Systems*, vol. 34, no. 2, pp. 597–626, Apr. 2017, doi: 10.1080/07421222.2017.1334499/suppl_file/mmis_a_1334499_sm1984.docx.
- [12] R. Wash and M. M. Cooper, “Who provides phishing training? Facts, stories, and people like me,” *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, doi: 10.1145/3173574.
- [13] S. Chaudhary, “The use of usable security and security education to fight phishing attacks,” Ph.D. Thesis, Nov. 2016, Accessed: Feb. 17, 2022. [Online]. Available: <https://trepo.tuni.fi/handle/10024/100073>
- [14] National Cyber Security Centre, “Cyber security for schools.” <https://www.ncsc.gov.uk/section/education-skills/cyber-security-schools> (accessed Feb. 17, 2022).
- [15] L. Jaeger and A. Eckhardt, “Eyes wide open: the role of situational information security awareness for security-related behaviour,” *Information Systems Journal*, vol. 31, no. 3, pp. 429–472, May 2021, doi: 10.1111/ISJ.12317.
- [16] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, “Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model,” *Decision Support Systems*, vol. 51, no. 3, pp. 576–586, Jun. 2011, doi: 10.1016/J.DSS.2011.03.002.
- [17] A. Vishwanath, “Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack,” *Journal of Computer-Mediated Communication*, vol. 20, no. 5, pp. 570–584, Sep. 2015, doi: 10.1111/JCC4.12126.
- [18] H. Abroshan, J. Devos, G. Poels, and E. Laermans, “A phishing mitigation solution using human behaviour and emotions that influence the success of phishing attacks,” *UMAP 2021 - Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pp. 345–350, Jun. 2021, doi: 10.1145/3450614.3464472.
- [19] N. A. G. Arachchilage and S. Love, “Security awareness of computer users: a phishing threat avoidance perspective,” *Computers in Human Behavior*, vol. 38, pp. 304–312, Sep. 2014, doi: 10.1016/J.CHB.2014.05.046.
- [20] W. He and Z. Zhang, “Enterprise cybersecurity training and awareness programs: recommendations for success,” *Journal of Organizational Computing and Electronic Commerce*, vol. 29, no. 4, pp. 249–257, Oct. 2019, doi: 10.1080/10919392.2019.1611528.
- [21] R. Wash, N. Nthala, and E. Rader, “Knowledge and capabilities that non-expert users bring to phishing detection,” *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*, 2021, pp. 377–396. Accessed: Feb. 18, 2022. [Online]. Available: <https://www.usenix.org/conference/soups2021/presentation/acar>
- [22] Witte, Kim (1992), “Putting the Fear Back into Fear Appeals: The Extended Parallel Process Model,” *Communication Monographs*, 59, 329–49. [Taylor & Francis Online], [Web of Science ®].
- [23] A. A. Hasegawa, N. Yamashita, M. Akiyama, and T. Mori, “Why they ignore english emails: the challenges of non-native speakers in identifying phishing emails”. *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*, 2021. Accessed: Feb. 18, 2022. [Online]. Available: <https://www.usenix.org/conference/soups2021/presentation/acar>
- [24] Gopavaram, Shakthidhar and Dev, Jayati and Grobler, Marthie and Kim, DongInn and Das, Sanchari and Camp, L. Jean, Cross-National Study on Phishing Resilience (May 7, 2021). In *Proceedings of the Workshop on Usable Security and Privacy (USEC)*, 2021, Available at SSRN: <https://ssrn.com/abstract=3859057>
- [25] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, “Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions,” *Proceedings of Conference on Human Factors in Computing Systems*, vol. 1, pp. 373–382, 2010, doi: 10.1145/1753326.1753383.

- [26] J. Sibrian, J. Mickens, and J. A. Paulson, “Sensitive data? now that’s a catch! The psychology of phishing,” Bachelor’s thesis, Jun. 2020, Accessed: Feb. 17, 2022. [Online]. Available: <https://dash.harvard.edu/handle/1/37364686>
- [27] N. Nthala and R. Wash, “How non-experts try to detect phishing scam emails”, In Workshop on Consumer Protection, Accessed: Feb. 17, 2022. [Online]. Available: <https://msucas-paid.sona-systems.com>
- [28] R. Wash, “How experts detect phishing scam emails,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, Oct. 2020, doi: 10.1145/3415231.

AI IN TELEMEDICINE: AN APPRAISAL ON DEEP LEARNING-BASED APPROACHES TO VIRTUAL DIAGNOSTIC SOLUTIONS (VDS)

Ozioma Collins Oguine and Kanyifeechukwu Jane Oguine

Department of Computer Science, University of Abuja, Nigeria

ABSTRACT

Advancements in Telemedicine as an approach to healthcare delivery have heralded a new dawn in modern Medicine. Its fast-paced development in our contemporary society is credence to the advances in Artificial Intelligence and Information Technology. This paper carries out a descriptive study to broadly explore AI's implementations in healthcare delivery with a more holistic view of the usability of various Telemedical Innovations in enhancing Virtual Diagnostic Solutions (VDS). This research further explores notable developments in Deep Learning model optimizations for Virtual Diagnostic Solutions. A further research review on the prospects of Virtual Diagnostic Solutions (VDS) and foreseeable challenges was also highlighted. Conclusively, this research gives a general overview of Artificial Intelligence in Telemedicine with a central focus on Deep Learning-based approaches to Virtual Diagnostic Solutions.

KEYWORDS

Biomedical imaging, Telemedicine, Smart Healthcare, Medical Imaging, AI, Virtual Diagnostic Solutions.

1. INTRODUCTION

Healthcare and Medicine are areas of modern society which has gained quite an outstanding level of research attention given current antecedents of virus outbreaks and spikes in anomalies regarding human health. Over the years, advancement in Artificial Intelligence and its resonating research areas such as Telecommunication and information technology has stirred up questions and advanced solutions regarding Human health. Affirmatively, we can infer that these improvements have notably impacted the medical and healthcare delivery scale and quality. However, healthcare access and delivery have struggled extensively to meet anticipated simultaneous prospects, as is the situation in many parts of the world, predominantly in underdeveloped and developing nations. A significant reason for this decline is the ever-increasing number of healthcare users and medical patients leading to the overutilization of medical resources such as healthcare providers, medical staff, and access to medical infrastructures. Hence, a more suitable and sustainable healthcare delivery approach was necessary to eliminate the issues arising from overpopulation and access to healthcare infrastructures.

A notable convergence of Machine Learning, Robotics, Telecom, computational neuroscience, and cloud computing has established a new infrastructure for global healthcare delivery known as *Telemedicine*. According to Khemapech et al., "Telemedicine is the delivery of health care services, with significant consideration of distance in service delivery and accessibility by all

stakeholders as key variables, using information and communications technologies” [1]. An inferred purpose of this innovation is to exchange valid information for diagnosing, treating, and preventing disease and injuries. Also, for research and evaluation, continuing education of health care providers and users, all aimed at advancing the healthcare systems of individuals and their communities. Although research on Telemedicine has been ongoing for decades, the emergence of the COVID-19 pandemic has reinvented its usage instantaneously, resulting in its scalability in fully implementing essential health and safety protocols in service delivery. This research area is an innovation that transcended from being a convenient alternative for technically savvy patients to the mainstay for healthcare delivery today across nations of the world, a reality that may continue long after the COVID pandemic. Telemedicine is the most suitable and sustainable approach to delivering healthcare services to patients by incorporating technological advancements in affordable and low-cost implementations. It also eliminates factors that adversely affect privileges to healthcare by strategically expanding virtual solutions to accommodate the growing populace and integrating smart tech to facilitate healthcare access and efficiency. The telemedicine framework has enabled collective improvements in Virtual Diagnostic Solutions, as shown in Fig 1.

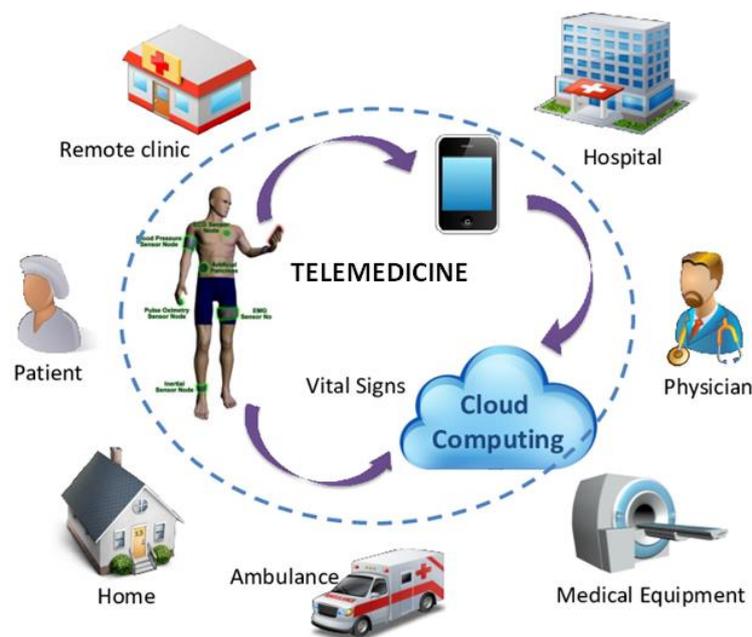


Fig 1. Telemedicine Infrastructure and Components

Deep Learning: Deep learning can be referred to as a modern state-of-the-art mechanism for computational processes applied to solving contemporary problems. Its potential for drawing patterns and insights from a massive amount of data boasts its wide adoption.

Virtual Diagnostic Solutions (VDS): This is a systematic approach to undertaking medical prognosis and administering or proffering medical solutions with little or no supervision from medical Professionals.

2. STAKEHOLDERS IN TELEMEDICINE

Telemedicine as an infrastructure for healthcare improvement entails the participation of four significant stakeholders, namely:

Patients (Key Stakeholders): These are the receivers or users of healthcare services. They are perhaps the most critical stakeholder to consider when thinking of approaches to Telemedicine.

Medical Professionals: These are the next most essential stakeholders in Telemedicine. Their knowledge area and research lay the foundation for building a telemedical Infrastructure.

Developers/ Tech Experts: These stakeholders are necessary because they constantly research innovative ways to build solutions to medical problems (Middlemen between Patients and Medical Professionals).

Policy Makers: These are folks who formulate and regulate laws that guide the implementation and utilization of Telemedicine

Telemedicine is currently gaining legislative and regulatory support in most developed countries; there have been no nationally representative estimates on its implementation by physicians and health practitioners across all medical specialties [2]. “To tackle this information gap, the American Medical Association (AMA) conducted a study in 2016 that surveyed 3,500 physicians who provided needed data to help assess potential barriers and create strategies to promote telemedicine adoption. The data analysis report from the AMA’s 2016 Physician Practice Benchmark Survey clearly described the Physician-to-Patient ratio of Virtual Diagnostic Solutions (VDS) in Telemedicine, as shown in Fig. 2. Consequently, this paradigm also proved efficient in maintaining and sharing patients’ medical records among hospitals and medical institutions.

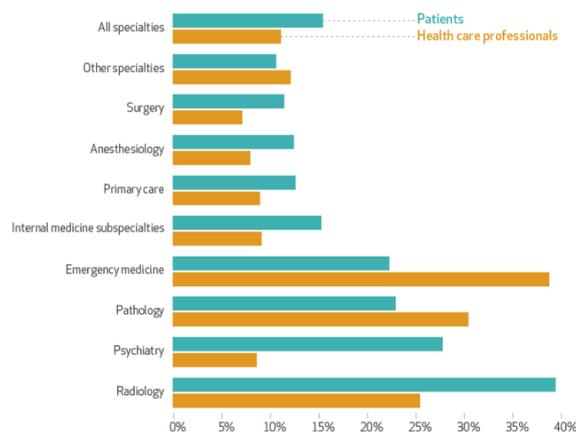


Fig 2. AMA 2016 Data Analysis on Patient to Healthcare Professional ratio in Telemedicine

The advancement of Artificial Intelligence and its research areas, such as the Internet of Things (IoT), Machine Learning, Image processing, and Deep Learning, has highly improved the level and quality of services provided by Telemedical software and applications. Researchers have and are currently up scaling on services and innovative solutions such as medical prognosis, Analytical Medicine, Robotic Surgery, DNA (genome) Sequence Analysis (DSA), Drug research and discovery, Medical Data Security, Clinical Trials, Medical Risk Prediction, emergency services, and Medical Image Analysis (brain monitoring, computer tomography, and radiology) offered by this remarkable means of healthcare delivery. Amid all the solution-oriented approaches put forward to dress health challenges, medical imaging is one area that has seen more promising results and prospects. X-Ray, Ultrasound (US), MRI, and CT Scanners have been interfaced with computers to transfer medical images to the remote center for careful analysis and early detection of medical abnormalities [3].

This paper explicitly discusses the implementation of artificial intelligence in Telemedicine with a more holistic view of Annotation-Efficient deep learning models for Medical Imaging in Virtual Diagnostic Solutions (VDS). It also elaborates on how data (input images) from medical imaging equipment are passed to sophisticated algorithms for accurate Diagnosis and treatment of Imaging-related ailments. Furthermore, application scenarios of medical Imaging models for VDS are subsequently highlighted and reviewed to understand its challenges and perceived prospects.

3. ARTIFICIAL INTELLIGENCE IN TELEMEDICINE

Artificial intelligence is a broad research field that facilitates machine simulation of human intelligence and behaviors ranging from learning to problem-solving. Its rapid growth and development in the past decade are due to its vast implementation in all human endeavors. Given the recent upsurge in big data generation, powerful computing coupled with refined computational models and algorithms, developments in AI have accelerated exponentially [4]. This trend has flagged the emergence of subfields such as Machine Learning (ML), Natural Language Processing (NLP), AI voice technology, Medical Imaging, AI assistants, Computer Vision, and robotics.

Telemedicine is a resurging innovation that has gained wide adoption in most developed and developing countries to Fastrack the accessibility to sustainable healthcare. It has achieved quite a laudable level of attention and research over the last two decades as ongoing studies are exploring ways to improve the existing infrastructure to a State-of-the-Art (SOTA). For efficient healthcare provision in contemporary societies, humans and machines have complimented each other in effectively delivering healthcare services through Virtual Diagnostic Solutions (VDS), thus providing a platform that bridges the gap between communication and accessibility to medical services. AI in Telemedicine has seen rapid adoption primarily based on the volume of medical data (Big data) recurrently generated. According to Ozioma et al., traditional data-handling techniques have proved ineffective in utilizing these data to obtain viable medical insights or solutions, given the gargantuan nature of the data [5]. Hence, sophisticated demonstrations of Artificial Intelligence approaches and models in Medical Diagnostics are becoming relatively popular. The adaptability and flexibility of these AI-based approaches and models have also driven the necessity for their implementation.

Andressa et al., amongst other scholars (see table 1), proposed an architecture that relies on fingerprinting and FLIPER framework to Fastrack the versatility and interconnectivity of healthcare applications. They also anticipated impact of their research was to enable quick, customizable resources that meet the level of reliability required for AI in Smart-health applications [6].

Table 1. Comparative Table describing Researches implementation of AI in Telemedicine [7]

Research Article	Year	Trend Category	Methodology
A Predictive Model for Assistive Technology Adoption for People with Dementia [8]	2014	Information Analysis and Collaboration	KNN and other data mining algorithms were utilized to analyze the behavior of patients with dementia and their adaptation to technology.

A Telerehabilitation Application with Pre-defined Consultation Classes [9]	2014	Healthcare Information Technology	Issues of Telemedicine under low-bandwidth network conditions were addressed using customized consultation classes demonstrating rehabilitation practices with preset parameters and a bandwidth adaption algorithm
An application of fuzzy systems was used to identify the best course of action for a given situation to Monitor Mobile Patients by Combining Clinical Observations with Data from Wearable Sensors [10]	2014	Intelligent Assistance Diagnosis	“This paper explores principled machine learning approaches to interpreting large quantities of continuously acquired, multivariate physiological data. Early warning of serious physiological determination was done using wearable patient monitors, such that a degree of predictive care may be provided.”
Ankle Rehabilitation System with Feedback from a Smartphone Wireless Gyroscope Platform and Machine Learning Classification [11]	2015	Patient Monitoring	This study uses a smartphone application, a wireless gyroscope platform, machine learning, and 3D printing to record usage and effects of therapy on an ankle and measure the strategy's efficacy.
Intelligent decision systems in Medicine -a short survey on medical Diagnosis and patient management [12]	2015	Intelligent Assistance Diagnosis	The study presents “a short review of some current Machine Learning algorithms (neural networks, genetic algorithms, support vector machines, Bayesian decision, k-nearest neighbor, etc.) used for automated Diagnosis of different major diseases, such as breast, pancreatic, and lung cancer, heart attacks, Diabetes.”
Smartphone-Based Recognition of States and Changes in Bipolar Disorder Patients [13]	2015	Intelligent Assistance Diagnosis	This paper proposes a system of using a smartphone-sensing wearable device to evaluate the behavior and recognize depressive and manic states of patients with bipolar disorder.
An Effective Telemedicine Security Using Wavelet-Based Watermarking [14]	2016	Information Technology	This paper proposes an algorithm that embeds and reads digital wavelet watermarks on medical images to secure confidentiality.
Mobile Cyber-Physical Systems for Health Care: Functions, Ambient Ontology and e-Diagnostics [15]	2016	Patient Monitoring	This paper proposes the use of a monitoring system embedded in wearable devices for the doctor or family members to receive updates on the patient's status.

Detection of Fetal Electrocardiogram through OFDM, Neuro-Fuzzy logic, and Wavelets Systems for Telemetry [16]	2016	Patient Monitoring	This study uses a neuro-fuzzy logic system to monitor and detect the exact electrocardiogram and other signals of a fetus inside an abdomen.
Using CART for Advanced Prediction of Asthma Attacks Based on Telemonitoring Data [17]	2016	Intelligent Assistance Diagnosis	This study created an algorithm with data from a home-based telemonitoring system to predict asthma exacerbation.
A Wireless Continuous Patient Monitoring System for Dengue: Wi-Mon [18]	2017	Patient Monitoring	The paper presents “a wireless monitoring system for patients who need continuous monitoring, using the Wireless Body Area Network (WBAN) concept.”

This research also aims to describe and evaluate the impact of Artificial Intelligence in Telemedicine. Some documented implementations of this powerful paradigm are discussed below:

Medical Imaging: This generally entails training AI models with images of medical scans that have been scientifically collected and stored in data repositories and databases. AI has significantly reduced the cost and time involved in analyzing scans through advanced deep learning models to diagnose health disorders. Hence, potentially allowing more scans to be taken and improving prognosis proficiency and accuracy [19]. A vast amount of resources and research has been expended in developing this field, as will be accentuated later in this paper. These researches have paved the way for State-of-the-art detection methodologies of medical ailments such as Brain tumors, skin and breast cancer, Pneumonia, and eye diseases.

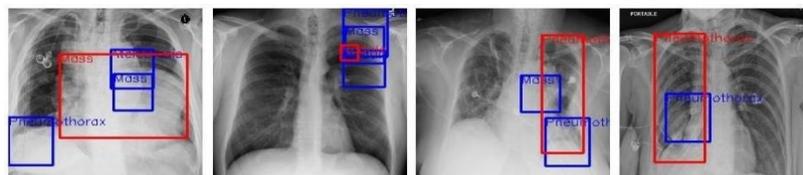


Fig 3. Covid-19 Detection in Lungs using Multibox SSD Model [20]

Echocardiography: Heartbeat patterns and coronary heart disease diagnosis and detection utilize the Ultramics system, an AI framework trialed at John Radcliffe Hospital in Oxford, to analyze echocardiography scans [21].

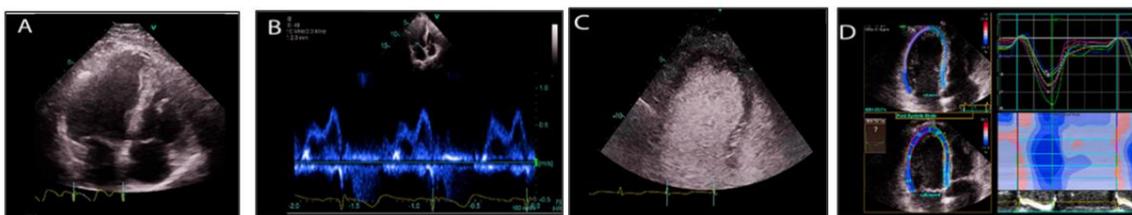


Fig 4. Classification of Echocardiograms using Deep Learning [22]

Screening for Neurological Conditions: Several AI models are being developed and employed in speech patterns analysis to predict psychotic episodes, Schizophrenia, recognize and monitor symptoms of neurological disorders such as Parkinson's disease [23].

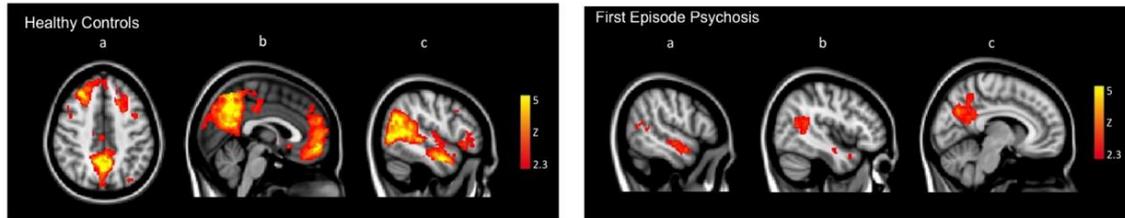


Fig 5. fMRI Study in Individuals with First-Episode Psychosis [24]

Emotion Recognition for Psychological Prognosis: Artificial Intelligence Deep learning models have proven efficient in recent times in observing and predicting emotions through individual reactions and psychological states at a given time. Ozioma et al. opined in their research on Facial Expression Recognition that Emotions are fundamental in human communication, driven by the erratic nature of the human mind and the perception of relayed information from the environment [25]. They proposed a hybrid deep learning model that makes real-time predictions of a person's emotional state, categorizing the individual's emotion into one of seven classes. The need for novelties in advancing this research field stems from the alarming rate of Emotion disorders, Suicide, Post-traumatic stress disorders (PTSD), and Psychological/Mental breakdown suffered by people in our contemporary society.



Fig 6. Facial Emotion Recognition using Hybrid Deep Learning Model [25]

Teleradiology: “Telecommunication is employed to transmit digital radiological images, like X-rays, Computed Tomograms (CTs), and Magnetic Resonance Images (MRI's) across geographical locations for interpretation and consultation” [26]. Reducing financial costs is one of the primary benefits of Teleradiology as it significantly reduces constraints in accessing radiological images, reports, and feedback between health professionals and patients.

Teleradiology is a crucial means for optimizing radiology workflow by sending the images to the radiologist rather than traditionally going to the radiology facility. By its very nature, Teleradiology is an efficient and high-quality manner by which patients' images can be interpreted and diagnosed by qualified specialists. Cloud services are primarily employed in this Telemedical service, where health stakeholders can utilize numerous privileges from the cloud, thus, upscaling the quality of radiology services. AI in Teleradiology will also enable the sharing of clinical information, medical imaging studies, and patient diagnostics [27] between patients and healthcare professionals.

Tele dermatology: This service implements Telecommunication to transmit medical information concerning skin conditions (e.g., tumors) for interpretation and consultation. According to Eedy and Wotton, “Tele dermatology model has received extensive advocate as a healthcare delivery model that may reduce inequalities encountered in the utilization of overstretched dermatological services, with implementation concentration ranging from remote to isolated communities” [28]. Landow et al., in their research, highlighted four factors that have stirred a relative increase in face-to-face appointments which tele dermatology strategies have tackled: (1) effective pre-selection of patients for teleconsultation, (2) high-quality photographic images, (3) dermoscopy if pigmented lesions are evaluated, and (4) adequate infrastructure and culture in place to implement teleconsultation recommendations [29].

4. APPLICATIONS OF DEEP LEARNING-BASED DIAGNOSTIC SOLUTIONS IN TELEMEDICINE

State-of-the-art Deep Learning models have seen advancements in methodologies over various medical problems such as object detection, recognition and segmentation in computer vision, voice recognition, and genotype/phenotype prediction. Telemedicine employs deep learning models in several Virtual Diagnostic Solutions today. While early studies focused on 2D medical images, such as chest X-rays, mammograms, and histopathological images, recent studies are looking toward applying sophisticated deep learning models to volumetric medical images.

CNNs form the basis for most State-of-the-art medical imaging DL models, which have gained wide prominence since achieving impressive results at the ImageNet [30] competition in 2012. Akkus et al. opined that CNNs remain a popular choice of DL approaches to image processing given their laudable tendency to weight sharing across convolutional layers or feature maps, in contrast, to fully connected ANNs [22]. And a rational reason for this was that, for 2D/3D image processing, ANNs utilize heavy computational processing that consumes a relatively high amount of GPU memory.

Several scholars have conducted evaluation studies and proposed several methodologies for different Virtual Diagnostic Solutions employing medical Imaging to solve health-related issues ranging from collecting image data to evaluating and diagnosing medical ailments. Qin et al. utilized Computer-aided Detection (CAD) in chest radiography based on contrast enhancement and segmentation in diagnosing various lung diseases such as early lung cancer, Pneumonia, Tuberculosis, and, more recently, lung inflammation levels caused by Covid-19 [31]. Numerous well-known DCNN architectures tested by Shin et al. emphasized the efficiency of transfer learning approaches in CT patch-based thoracoabdominal lymph node detection and ILD classification [32]. A Recurrent full Convolutional Neural Network (RFCNN) proposed by Poudel et al. was used to segment the left ventricle from cardiac MR images [33]. Hosseini-Asl et al. proposed a 3D CNN model to determine the progression of Alzheimer’s disease from structural brain MR images [34]. In the method, they employed a transfer learning approach that utilized pre-trained weights of features from a Computer-Aided Engineering (CAE) with a small number of source domain images to fine-tune the target domain data to train the actual classification model. Yu et al. proposed a 3D volumetric CNN for prostate segmentation on MR images through U-net expansion, initially used for 2D biomedical image segmentation. They also added residual connections to combine multiple-scale information [35, 36]. An alternative study proposed a slice-level classification model to detect Interstitial Lung Diseases (ILD) from chest CT scans [37]. Jamaludin et al. detected several diseases simultaneously from spinal MR images through a trained multi-task learning model. They visualized salient regions in the image for corresponding predictions as ‘evidence hotspots’ as seen in Fig.7 [38].

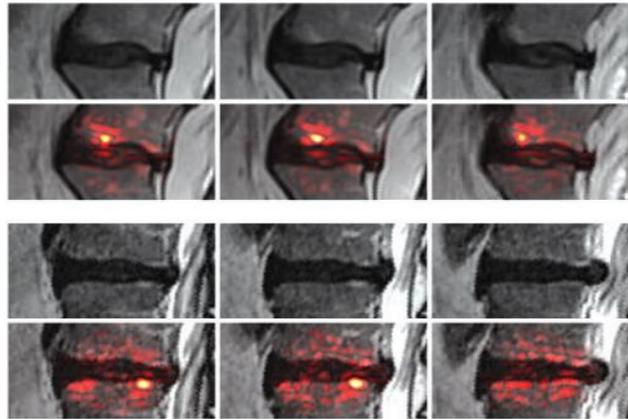


Fig 7. “Evidence hotspot” Visualization in Spinal MRI [38]

5. PROSPECTS OF AI-BASED VIRTUAL DIAGNOSTIC SOLUTIONS (VDS) IN TELEMEDICINE

Early Diagnosis and treatment of diseases have been a primary focus of modern-day healthcare infrastructure. In previous sections above, this research has significantly highlighted some outstanding advancements and implementation of the Deep Learning approach of Artificial Intelligence to improve Virtual Diagnostic Solutions. This review paper also seeks to forecast based on the current state-of-the-art prospects of this frequently evolving paradigm. A crucial benefit of this paper is to buttress the promises of AI in future healthcare development.

Early Diagnosis and Treatment: Diagnosis and treatment have ensured the advancement of research conducted in the field of Medicine today. With the current pace in development, a significant prospect of AI in Telemedicine is seen to be geared towards establishing and improving early diagnosis and treatment mechanisms of numerous medical conditions. Jacobsmeier showed the effectiveness of AI in the early detection and management of infectious diseases and epidemics such as Water contamination, worldwide virus outbreaks, etc. [39]. Research Labs and tech firms are currently pushing the limits on a proactive scale in ensuring this, as not a week goes by without the introduction of new approaches to medical solutions using AI or Big data with improved accuracy and precision.

AI systems will also become more advanced in engaging in broader medical problem-solving tasks with little human supervision or control. Hence, advanced ethical learning as imposed by state-of-the-art deep learning models will play a crucial role in the effective implementation strategy and adoption of Artificial Intelligence approaches in Virtual Diagnostic Solutions.

Independence: A major gap identified in traditional Medicine is the inability of patients to access healthcare when, where, and how they want it. Deep learning approaches to Telemedicine have gone far beyond filling this gap by empowering patients to systematically, thematically, and objectively evaluate symptoms and proffering possible solutions through virtual diagnostic solutions. Another significant advantage is the noteworthy reduction in the strain on medical resources and healthcare professionals. While concerns have been raised regarding the negative impact of AI solutions on Telemedicine [19], several notable research has been observed to acknowledge its importance and necessity in delivering cost-effective, high quality and accessible healthcare services [40, 41].

Improved Clinical and Therapeutic Coordination: Not only have the efficiency and advancements of AI ensured quality in the level of medical potential, but they have also shown a remarkable promise in the level of service delivery mechanisms. As more implementations of Virtual Diagnostic Solutions are utilized, huge amounts of data are also generated from healthcare professionals and patients. Over time, this aggregate amount of data will create an abundance of information on stakeholders involved in Telemedicine utilization. A foreseeable impact of this will be the development of systematic and coordinated clinical therapeutic processes and services that ensure that VDS prognosis are subjected to second opinion evaluations to establish accurate Diagnosis, the improved continuity in medical care through research and training mechanisms, simplification of medical procedures and the analysis of medical inconsistencies and patterns in patient health to foster the development of new patient-centered models.

Scalability: Global healthcare witnessed new dawn with the discovery of Telemedicine which ensured that medical solutions could be provided outside hospital buildings. Virtual workflows and technologies have been set up over time to create optimal infrastructures to implement this process. The scalability of Virtual Diagnostics Solutions used in Telemedicine has been a significant metric of development in our contemporary society. Evolutionary standards in AI have ultimately ensured the continuous upgrade of Telecommunications, software, and digital technologies used to implement virtual and real-time diagnostic solutions. A relative expansion in the number of telemedical services provided by Virtual Diagnostic Solutions has spiraled the frequent adoption of this paradigm in tackling global healthcare challenges such as epidemics, pandemics, and critical health conditions.

Another reason for improvements in scalability is the ever-changing needs of healthcare stakeholders to meet state-of-the-art next-generation service level requirements. Scalability could entail efficient resource allocation, enhanced data integrated hospital-grade wearable devices, expansion of patient management software systems, robust databases, cloud services, and data-sharing infrastructures. As we advance, a perceived rise in the adoption of the telemedical model of healthcare delivery will no doubt facilitate accessibility and efficiency with the sole purpose of bringing a paradigm shift to Global health solutions.

6. CHALLENGES OF AI-BASED VIRTUAL DIAGNOSTIC SOLUTIONS (VDS) IN TELEMEDICINE

Despite the numerous benefits promised by the integration of AI in Virtual Diagnostics Solutions (VDS) to provide geographically accessible, affordable, acceptable, and quality healthcare, there have also been challenges mitigating its full-scale adoption and development potential. As stated earlier, this paper will highlight some key barriers facing the advancement of Virtual Diagnostic Solutions in Telemedicine.

High-Cost: Although Virtual Diagnostic Solutions have been leveraged to reduce the high cost of healthcare accessibility, this is only valid from a patient's perspective. While this issue might seem improbable in our contemporary society, it is pertinent to note that '*Cost*' sums up the general resources required to develop, implement, maintain and advance Virtual Diagnostic Solutions in Telemedicine. These resources include but are not limited to human, technical, financial, and academic resources that have become relatively expensive to acquire and employ over time. Cost implications for efficient initiation and delivery of telemedical research and projects are seen as economic excess and, as such, given less consideration. This effect has led to the relatively reduced advocacy for adopting telemedical solutions in most developing and underdeveloped countries. From an AI perspective, huge amounts of technical resources,

skillsets, and training are required to build robust Telemedical deep learning models for Virtual Diagnostic Solutions. Financial resources are expended on the acquisition, installation, utilization, and maintenance of telemedical equipment. On a more holistic scale, this factor is the most critical, responsible for the decline and a foreseeable reduction in VDS development and adoption in Telemedicine.

Unavailability and Underdevelopment of Technical Infrastructures: Technology, Telecommunication, and Artificial Intelligence advancements are the backbone for functioning state-of-the-art Telemedical solutions, as have been elucidated in earlier sections of this paper. Hence, the robust nature of Telemedicine has necessitated the requirement of sophisticated infrastructures to ensure efficient and effective development and deployment. For instance, Teleophthalmology, Teleradiography, and real-time emergency consultation, amongst others, are some of the telemedical services which require heavy computing technologies and fast internet connectivity. Systematically excellent and diverse medical data are also needed for training Machine Learning (ML) and deep learning models to enhance effective generalization. Technology literacy levels also play a part in ensuring telemedical services' smooth implementation and sustainability by abstracting the process workflow to patients and medical professionals. The unavailability of these requirements is predicted to pose a significant gap in harnessing the potential of Telemedicine. On a more holistic scale, a lack of technical infrastructures will more likely hinder the development of Telemedicine, thereby causing a decline in the progress achieved so far. Despite the encouragement by the WHO encouraging the adoption of Telemedical innovations by member states, a major drawback emanating from this challenge has been observed in underdeveloped and developing countries.

Reliability Issues: A significant concern regarding the adoption and implementation of Virtual Diagnostics Solution is its reliability and generalization ability. Given this issue's validity, several Virtual Diagnostics solutions have come under severe scrutiny, raising questions and sentiments concerning their utilization. Reliability is a dominant parameter in determining the rate of adoption and research in Telemedical solutions. Hence, at any slightest detection of untrustworthiness or inconsistency, Telemedicine could lose the attention and participation of key stakeholders. Error-prone VDS solutions have stereotypically discouraged total reliability in the potentials of Telemedicine. While a significant level of this challenge is due to lapses in technicalities and operational models, several other factors can also contribute to this issue, such as stakeholders' resistance to accepting change, Digital illiteracy leading to poor awareness of modern tech, cultural perceptions, and malpractice liabilities.

Ethical Violations, Confidentiality, and Privacy issues: Ever since the revolutionization of the Internet, privacy and ethical policy stability have been a critical challenge among internet users. Just like Medical professionals, patients (VDS users) need orientation and training on data access, privacy and protection measures when utilizing services on Virtual Diagnostic Solutions. For efficiency of Telemedicine, data privacy such as Doctor-Patient confidentiality, training, and licensing of personnel is expected to be strictly adhered to. However, stability in providing robust security infrastructures and interoperability features to tackle these challenges, coupled with the ever-rising trend of cyber-crimes, has hindered the trustworthy adoption of Telemedicine as an effective approach to healthcare access. Another propelling reason for this challenge is the result of insufficient legal policies, guidelines, and Standard Operating Procedures (SOPs). In addition to the absence of defined policies and regulatory procedures, a lack of international regulatory uniformity has stirred several controversies regarding Telemedical services and solutions. While AI's propensity for good has been established in earlier sections of this paper, Malpractice liability is another factor to consider. This issue has necessitated holistic reflections on AI's dual potentials by governments, Researchers, and Engineers developing Telemedical solutions.

Continuity/Sustainability: The recent global pandemic has necessitated the sustainability of advanced medical technologies. As more medical issues arise, so should the pervasiveness of healthcare approaches employed for treatment purposes. A major drawback, especially in underdeveloped and developing countries, is their incapability to adequately sustain and encourage advancements in Telemedicine either through research or system analysis. Several factors can be attributed to this enthusiastic acceptance of the already existing Virtual Diagnostic Solutions, formulation of ethical and privacy policies, standardization of technological equipment, skill set and infrastructures, cost-effectiveness and coordination, etc. In light of this, the sustainability and continuity of this ever-growing trend hinge on the improvements of all participating stakeholders in creating stable infrastructures and approaches to tackle existing challenges and foster the growth and application of AI in Virtual Diagnostic Solutions (VDS).

7. CONCLUSION

Artificial Intelligence has no small impact on Global health in our contemporary society. In most developed and some developing countries, Telemedicine has gained popularity for its benefits in improving healthcare access, reducing healthcare costs, and enhancing the quality of healthcare services. These benefits necessitated the rapid paradigm shift from the usual traditional healthcare (provider-centric) infrastructures to more robust (patient-centric) methodologies and infrastructures, given the tremendous pressure on healthcare providers to provide affordable, accessible, and quality healthcare services. This paper holistically discussed the growth and revolution of Telemedicine as a modern research field; it also introduced a broad insight towards implementing Artificial Intelligence in Telemedicine with specific inclinations toward services provided by Virtual Diagnostic Solutions (VDS). A review of works of literature from researchers citing several Deep Learning Approaches employed in detecting and treating several medical ailments was also discussed in this paper, recommending the importance of deep learning in the advancement of AI services. Several Applications of Artificial Intelligence in sustainable healthcare provision and access were also reviewed and described.

The overall significance of this paper is to throw more light on the importance of **DEEP LEARNING-BASED** AI methodologies in advancing state-of-the-art Virtual Diagnostic Solutions in Telemedicine. Consequently, this paper enumerated and discussed crucial prospects and challenges associated with incorporating, implementing, adopting, and advancing Artificial Intelligence in Telemedicine. Further holistic research on advanced Telemedical approaches is encouraged in tandem with these views.

ABBREVIATIONS

IoT - Internet of Things	CAD - Computer-Aided Detection
DSA - DNA Sequence Analysis	AI - Artificial Intelligence
VDS - Virtual Diagnostic Solutions	MRI - Magnetic Resonance Imaging
ML - Machine Learning	US - Ultrasound
CT - Computed Tomography	AMA - American Medical Association
NLP - Natural Language Processing	SOTA - State-of-the-Art
CAE - Computer-Aided Engineering Network	RFCNN - Recurrent Fully Convolutional Neural Network
ILD - Interstitial Lung Diseases	SOPs - Standard Operating Procedures

DISCLOSURE

The authors declare that they have no competing interests with anyone in publishing this paper.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions in conscripting the paper and revising it critically for important intellectual content; agreed to submit it to the current journal; and final approval of the version.

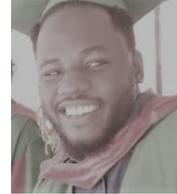
REFERENCES

- [1] T. Khemapech, W. Sansrimahachai, and M. Toahchoodee, "Telemedicine—Meaning, Challenges, and Opportunities," *Siriraj Medical Journal*, 2022. [Online]. Available: <http://dx.doi.org/10.33192/Smj.2019.38>. [Accessed: 06- Jul- 2022].
- [2] C. Kane, "AMA offers first national estimate of telemedicine use by physicians," *American Medical Association*, 2018. [Online]. Available: <https://www.ama-assn.org/press-center/press-releases/ama-offers-first-national-estimate-telemedicine-use-physicians>. [Accessed: 12- Jun- 2022].
- [3] Andr'e Pereira Alves, Tiago Marques Godinho, and Carlos Costa, "Assessing the relational database model for optimization of content discovery services in medical imaging repositories," *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1-6, 2016.
- [4] M. Chen and M. Decary, "Artificial intelligence in healthcare: An essential guide for health leaders," *Healthcare Management Forum*, vol. 33, no. 1, pp. 10-18, 2019. Available: <https://doi.org/10.1177/0840470419873123>. [Accessed 2 June 2022].
- [5] O. Oguine, K. Oguine, and H. Bisallah, "Big Data and Analytics Implementation in Tertiary Institutions to Predict Students Performance in Nigeria," *Science Open*, 2021. Available: [10.14293/s2199-1006.1.sor-ppfhsfb.v1](https://doi.org/10.14293/s2199-1006.1.sor-ppfhsfb.v1) [Accessed 21 June 2022].
- [6] A. Banerjee, C. Chakraborty, A. Kumar, and D. Biswas, *Emerging trends in IoT and big data analytics for biomedical and health care technologies*. Elsevier Inc., 2019.
- [7] D. M. Mitch, E. D. Subido, and N. T. Bugtai. "Trends in telemedicine utilizing artificial intelligence" *AIP Conference Proceedings* 1933, 040009 (2018); <https://doi.org/10.1063/1.5023979> Published Online: 13 February 2018.
- [8] Shuai Zhang, S. I. McClean, C. D. Nugent, M. P. Donnelly, L. Galway, B. W. Scotney, and I. Cleland, *IEEE J. Biomed. Heal. Informatics* 18(1), pp. 375–383 (2014).
- [9] T. K. Kiong and A. S. Narayanan, "A telerehabilitation application with pre-defined consultation classes," in *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*, (2014, pp. 1238–1244.
- [10] L. Clifton, D. A. Clifton, M. A. F. Pimentel, P. J. Watkinson, and L. Tarassenko, *IEEE J. Biomed. Heal. Informatics*, 18(3), pp. 722–730(2014).
- [11] R. LeMoyné, T. Mastroianni, A. Hessel, and K. Nishikawa, "Ankle Rehabilitation System with Feedback from a Smartphone Wireless Gyroscope Platform and Machine Learning Classification," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, (IEEE, Miami, FL, 2015), pp. 406–409.
- [12] F. Gorunescu, "Intelligent decision systems in Medicine? A short survey on medical diagnosis and patient management," in *2015 E-Health and Bioengineering Conference (EHB)*, (IEEE, Iasi, 2015), pp. 1–9.
- [13] A. Grunerbl, A. Muaremi, V. Osmani, G. Bahle, S. Öhler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz, *IEEE J. Biomed. Heal. Informatics* 19(1), pp. 140–148 (2015).
- [14] J. Singh and A. K. Patel, "An effective telemedicine security using wavelet-based watermarking," in *2016 IEEE (ICCIC)*, (IEEE, Chennai, 2016), pp. 1–6.
- [15] A. Costanzo, A. Faro, D. Giordano, and C. Pino, "Mobile cyber-physical systems for health care: Functions, ambientontology, and e-diagnostics," in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, (Las Vegas, NV, 2016) pp. 972–975.
- [16] P. Kumar, S. K. Sharma, and S. Prasad, "Detection of fetal electrocardiogram through OFDM, neuro-fuzzy logic and wavelets systems for telemetry," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, (IEEE, Coimbatore, 2016), pp. 1–4.
- [17] J. Finkelstein and I. C. Jeong, "Using CART for advanced prediction of asthma attacks based on telemonitoring data," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, (IEEE, New York, NY, 2016), pp. 1–5

- [18] S. Nubenthan and C. Shalomy, "A wireless continuous patient monitoring system for dengue: Wi-Mon," in 2017 6th National Conference on Technology and Management (NCTM), (IEEE, Nagercoil, 2017), pp. 23–27.
- [19] House of Lords Select Committee on Artificial Intelligence (2018) AI in the UK: ready, willing, and able?
- [20] F. Saiz, and I. Barandiaran, "COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach". International Journal of Interactive Multimedia and Artificial Intelligence. In Press. 1. 10.9781/ijimai.2020.04.003, 2020.
- [21] See <http://www.ultromics.com/>
- [22] Z. Akkus et al., "Artificial Intelligence (AI)-Empowered Echocardiography Interpretation: A State-of-the-Art Review." *J. Clin. Med.* 2021, 10, 1391. <https://doi.org/10.3390/jcm10071391>
- [23] G. Bedi et al., "Automated analysis of free speech predicts psychosis onset in high-risk youths, NPJ Schizophrenia, 1: 15030; IBM Research (5 January 2017) IBM 5 in 5": with AI, our words will be a window into our mental health.
- [24] C. Bartholomeusz et al., "An fMRI study of the theory of mind in individuals with first-episode psychosis," *Psychiatry Research: Neuroimaging*, vol. 281, pp. 1-11, 2018. Available: 10.1016/j.pscychresns.2018.08.011 [Accessed 26 June 2022].
- [25] O. C. Oguine, K. A. Kinfu, K. J. Oguine, H. I. Bisallah and D. Ofuani, "Hybrid Facial Expression Recognition (FER2013) Model for Real-Time Emotion Classification and Prediction", arXiv.org, 2022. [Online]. Available: <https://doi.org/10.48550/arXiv.2206.09509>. [Accessed: 22- Jun- 2022].
- [26] M. Fatehi, R. Safdari, M. Ghazisaeidi, M. Jebraeily, and M. Koolae, "Data Standard for in Tele-radiology.", *Acta Inform. Med.* 23(3): 165- 168, 2015.
- [27] E. J. Monteiro, C. Costa, and J. L. Oliveira, "A Cloud Architecture for Teleradiology-as-a-Service". *Methods. Inf. Med.* 55(3):203-14, 2016.
- [28] D. J. Eedy, and R. Wootton, "Teledermatology: a review." *Br. J. Dermatol.* 144:696-707, 2001.
- [29] S. M. Landow, A. Mateus, K. Korgavkar, D. Nightingale, and M. A. Weinstock, "Teledermatology: key factors associated with reducing face-to-face dermatology visits." *J. Am. Acad. Dermatol.* 71(3):570-6, 2014.
- [30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 2015, 115, 211–252. [CrossRef].
- [31] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey." *Biomed Eng Online* 17(1):113, 2018.
- [32] H. C. Shin, H. R. Roth, M. Gao, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics, and Transfer Learning." *IEEE Trans Med Imaging* 2016; 35:1285–98.
- [33] R. P. K. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation." Vol. 10129 of *Lecture Notes Comput Sci* 2017:83–94
- [34] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network", Arxiv.org, 2016. [Online]. Available: <https://arxiv.org/pdf/1607.00455.pdf>. [Accessed: 20- Jun- 2022].
- [35] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, "Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images". *Thirty-First AAAI Conference on Artificial Intelligence* 2017; 66–72.
- [36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." *MICCAI* 2015;234–41.
- [37] M. Gao, Z. Xu, L. Lu, A. P. Harrison, R. M. Summers, and D. J. Mollura, "Holistic Interstitial Lung Disease Detection using Deep Convolutional Neural Networks: Multi-label Learning and Unordered Pooling." arXiv [Internet] 2017;9352:1–9. Available from: <http://arxiv.org/abs/1701.05616/>.
- [38] A. Jamaludin, T. Kadir, and A. Zisserma, "SpineNet: Automatically pinpointing classification evidence in spinal MRIs." Vol. 9901 of *Lecture Notes Comput Sci* 2016:166-75.
- [39] B. Jacobsmeier, Focus: tracking down an epidemic's source *Physics* 5: 89; 2012.
- [40] World Health Organization, and WHO Telemedicine, "Opportunities and Developments in Member States, Reports on the Second Global Survey on eHealth 2009", *Global Observatory for eHealth Series, Volume 2*, 2010
- [41] Q. A. Qureshi, I. Ahmad, and A. Nawaz, "Readiness for E-Health in the Developing Countries Like Pakistan," *Gomal Journal of Medical Sciences*, Volume 10, No.1, Peshawar, Pakistan, 2012.

AUTHORS

Ozioma Collins Oguine is a Graduate Research Assistant at the University of Abuja, Nigeria. He graduated from the same University with First Class Honors (Summa cum Laude), top 1% from the Department of Computer Science. His research interests are Machine/Deep Learning, Computer Vision, Robotics, and Human-Computer Interaction (HCI). He is a Member of the Intelligent Automation Network (IAN), Black in AI, Black in Robotics, an illustrious member of the International Society of Engineers (IAENG) in Artificial Intelligence and Computer science, and an Associate Member of the British Computing Society (BCS).



Kanyifechukwu Jane Oguine is a Graduate Research Assistant at the University of Abuja, Nigeria. She graduated from the same University with First Class Honors (Summa cum Laude), top 2% from the Department of Computer Science. Her research interests are Machine/Deep Learning, Computer Vision, Computational Algorithm, and Human-Computer Interaction (HCI). She is a Member of the Intelligent Automation Network (IAN), Black in Robotics, also a notable member of the International Society of Engineers (IAENG) in Artificial Intelligence and Computer science, and an Associate Member of the British Computing Society (BCS).



AUTHOR INDEX

<i>Ajai Kumar</i>	55
<i>Amjad Alghamdi</i>	185
<i>Arwa Alsahli</i>	185
<i>Asrar Almogbil</i>	185
<i>Ciaran Haines</i>	69
<i>Damodar M</i>	55
<i>Deng Yang</i>	89
<i>Fadiyah Alghamdi</i>	185
<i>Farshad Ahmadi As</i>	131
<i>Feng Yuanyuan</i>	89
<i>Hans Li</i>	141
<i>Hisham Ihshaish</i>	69
<i>Igor Borovikov</i>	01
<i>J. Ignacio Deza</i>	69
<i>Jacob Hauenstein</i>	207
<i>Jawaher Alotaibi</i>	185
<i>Jiayi Zhang</i>	113
<i>Jiayu Zhang</i>	113
<i>John Jenq</i>	175
<i>John Zhang</i>	121
<i>Jon Rein</i>	01
<i>Justin Wang</i>	113
<i>Kai Segimoto</i>	101
<i>Kanyifeechukwu Jane Oguine</i>	229
<i>Karine Levonyan</i>	01
<i>Khuzama Ammar</i>	25
<i>Kinan Mansour</i>	25
<i>Kishor Patil</i>	55
<i>Li Weihao</i>	89
<i>Liu Kejian</i>	89
<i>Maan Ammar</i>	25
<i>Mehmet Bodur</i>	131
<i>Mousa Jari</i>	217
<i>Mozhdeh Sarkhoshi</i>	17
<i>Muhammad Yasir Adnan</i>	197
<i>Neha Gupta</i>	55
<i>Nelly Segimoto</i>	101
<i>Nitish Victor</i>	01
<i>Ozioma Collins Oguine</i>	229
<i>Pawel Wrotek</i>	01
<i>Peter Mayhew</i>	69
<i>Qianmu Li</i>	17
<i>Qingbo Wang</i>	45
<i>Qingshan Shen</i>	45
<i>Razan Alajlan</i>	185

<i>Richard Self</i>	197
<i>Ryan Fellows</i>	69
<i>Sagina Athikka</i>	175
<i>Sascha Alpers</i>	169
<i>Steve Battle</i>	69
<i>Waad Ammar</i>	25
<i>Yang Cheng</i>	89
<i>Yang Pachankis</i>	151
<i>Yong Xue</i>	197
<i>Yu Sun</i>	101,113,121,141